# MINIMAL ROUTE SEARCH BY USING SPATIAL QUERY KEYWORD

**S.RADHIKA[*1] B.ARAVIND REDDY[*2]**
[*1]Assistant Professor, Department of Computer Science
[*1]Nalla Narsimha Reddy educational society's Group of institutions
[*2]Senior Quality Analyst, IBM

***Abstract* – Minimal route search technique by using spatial query keyword. It operates on spatial data objects stored in database and using algorithms that can retrieve results very quickly. Best keyword cover query target is to find keywords associated with instances of particular class. In this paper we proposed a technique for optimal route search which considers the keywords like keyword relevance keyword rating and spatial relevance.**

***Keywords – spatial objects, Spatial keyword query, spatial database, Boolean range query, best keyword cover query.***

## I. INTRODUCTION

Data mining extracts the hidden patterns from a large datasets for researchers to use it for various purposes like decision making. Such kind of data plays a major role in keyword searching. Extracting features of spatial objects from spatial database has increased interest among the researchers. This motivated them to develop methods to retrieve a spatial object that consists of objects associated with spatial features.A spatial object includes longitude and latitude of location along with spatial data. Querying such kind of data is called as best keyword cover querying. Standard K-nearest neighbor (KNN) search and compute nearest neighbor by considering distance as feature or vector. In this context, nearest neighbor search focus on finding nearest neighbors with the help of keywords and spatial data. Hence these keywords play a major role in finding a minimal route.

## II. PROBLEM DEFINITION

The content used for querying will be in the form of spatial database. For example, consider a college given a spatial database P, which consists some set of points. For a query q, where q belongs to set of objects, initially it search within the cluster of that object belongs and then it performs nearest neighbor search to obtain the result to that query. The idea of keyword rating was introduced along with its characteristics other than distance for such search, query will take the form of features of objects for better decision making. It finds the nearest neighbour based on a similarity measure, like weighted average of index rating which includes keyword rating, keyword search and nearest neighbor search.

## II. PROPOSED METHOD

### A. Baseline Algorithm

Baseline algorithm mainly focuses on retrieving the data with respect to query keywords. Best keyword cover can be obtained by this algorithm which requires spatial objects to be in the form of files which contains features like location, document identifier and its address. All operations will be revolving around the spatial objects. These objects are the Inputs to baseline algorithm for single query keyword. The initial step of this algorithm is to set a variable as zero. Next it generates candidate keyword cover. The Candidate keyword cover will generate spatial objects that contain the query keywords. Significance of Keyword will be calculated using term frequency inverse document as similarity measure. Term frequency is measured by the number of times that a term occurs in a single document divided by the total number of terms available in a document. Inverse document frequency is measured by the inverse number of times that a term occurs in a document divided by the total number of documents available. TFIDF is the Significance of Keyword will be calculated using term frequency inverse document as similarity measure. Term frequency is measured by the number of times that a term occurs in a single document divided by the total number of terms available in a document. Inverse document frequency is measured by the inverse number of times that a term occurs in a document divided by the total number of documents available. TFIDF is the product of both term frequency and inverse document frequency. Initially the default value is set to zero and the newly obtained score is compared with the first score. If that score is greater than zero, then will be set as best keyword cover. The next step is to perform a nearest neighbor search using a traditional similarity measure named Euclidean distance(ED). Nearest neighbor search algorithm sets its default value of users as current user location. Based on that location, rest of the distance will be calculated with respect to that location. The one least distance with respect to query location will be considered as best keyword covers. But as number of query keywords increases, its performance gradually drops and running Time will be very high.

## IV. EXPERIMENT

Baseline algorithm applied on this real time data target to extract data using single query keyword. Keyword nearest neighbor (NN) expansion variant algorithm retrieves data using multiple query keywords.

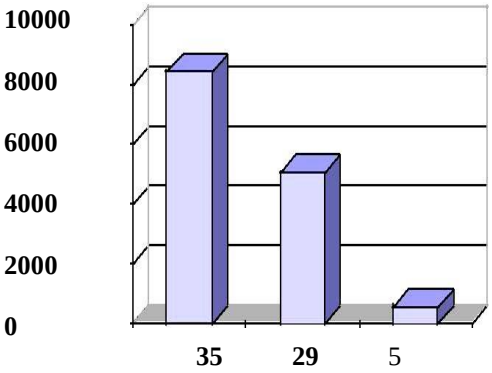


*Fig. 1 Execution time versus dataset count of baseline algorithm*

Fig 1 shows a bar chart representing execution time of baseline algorithm versus dataset count or number of files searched for a particular query keyword. Vertical axis indicates execution time during search process. When search procedure occurs Execution time is the difference between start time and run time. When a single query keyword is searched in a file count of thirty five, its execution time is 8533 milliseconds. When searching takes place in a file count of twenty nine, execution time is 5091 milliseconds similarly, when searching takes in a file count of five for one query keyword, its execution time is 708 milliseconds. It has been observed that execution time increases as files to be searched increases. Hence File count is linearly proportional to execution time.

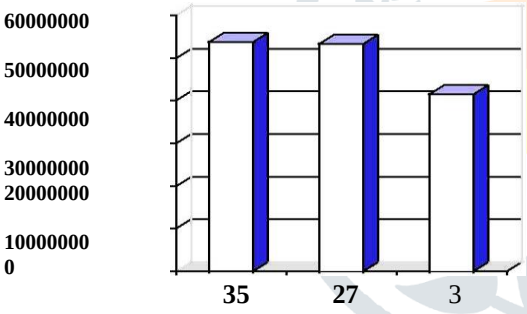


Fig 2 shows memory consumption with respect to the file search of baseline graph. Where X axis indicates memory consumption time and Y axis indicates the number of files or dataset count. When file searched is thirty five, memory required is 52376675 KB. When file searched is three for a particular query keyword, memory required is 42422105 KB. When file searched is twenty seven for a particular query keyword, memory required is 52161130 KB.. From this graph, it has been shown that as number of files searched increases memory required increases. Hence the number of files searched is directly proportional to memory consumption.

## V. CONCLUSION

A comprehensive report of two algorithms to retrieve best keyword cover is applied on real data with two dimensions. Best keyword covers query targets to recover spatial objects with respect to user's requirement. Algorithms are used to find answers to such queries. This can be used with best keyword cover route search which helps to find best route.

## REFERENCES

[1] Ke Deng, Xin Li and Xiaofang Zhou, "Best Keyword Cover Search", IEEE Transaction on Knowledge and Data Engineering ,vol 27, no 1, January 2015.

[2] Joao B Junior, Orestis Gkorghas, Simon Jonassen and Kjetil Norvag, " Efficient Processing of Top K Spatial Keyword Queries", in SSTD, pages 205-222, 2012.

[3] Xing Cao, Gao Cong, Beng Chin, "Collective Spatial Keyword Querying", ACM Transaction on Database Systems, 2011.

[4] Gils R Haltom and Hanseb Samet, "Distance Browsing in Spatial Databases", ACM Transaction on Database Systems, June 1999, pp 265-318.

[5] Ronald Fagin, Amman Lotem and Moni Naor, "Optimal Aggregation Algorithms for Middleware", Journal of Computer and System Sciences, April 2003.

[6] Yulee Tao and Cheng Sheng, " Fast Nearest Neighbor Search with Keywords", IEEE Transaction on Knowledge and Data Engineering, vol 26, no 4, April 2014.

[7] Lisi Chen, Geo Cong, Christian S Jensen, "Spatial Keyword Query Processing: An experimental evaluation", Proceedings of the VLDB Endowment, vol 6, no 3, 2013

[8] DongXiang Zhang, Kian Lee Tan, Anthony K. H. Tung,, "Scalable Top K Spatial Keyword Search", ICDT , March 2015.

[9] Bolin Ding, Bo Zhao, " Efficient Keyword Based Search for Top K Cells in Text Cube", IEEE Transaction, December 2011.