



## Finding Nearest Neighbour Search Using Spatio-Text Inverter

**T. Suganiya**

M.E-CSE&NPR College of  
Engineering and Technology,  
Tamilnadu, India

**M. Akila Rani**

Assistant Professor & NPR College  
of Engineering and Technology,  
Tamilnadu, India

---

**Abstract**— *Spatio-textual queries recover the greatest related items with respect to a given location and keyword set. It gives more attention based on how to find top-k result due to spatio-textual query. Users don't know the exact keywords and they choose them from candidate keyword set. The system introduces a novel method such as spatio-text inverter method and multiple query processing method. The multiple queries are very useful to user side because they give more related details and get accurately by using top-k search methods. It will improve the effectiveness and efficiency, then it focuses on the top-k matches for the query due to k values.*

**Keywords**— *Spatio-text inverter, Multiple query processing, top-k search, Keyword set, Spatio-text query*

---

### I. INTRODUCTION

Geo Spatial is data or information that identifies the geographic location of features and boundaries on earth, such as natural or constructed features, oceans, and more. It is an approach to applying statistical analysis and other informational techniques to data which has a geographical or geographical aspect. Such analysis would typically employ software capable of geospatial representation and processing, and apply analytical methods to geographic datasets.

Spatio-textual query that takes a user location and a keyword set as inputs, and returns the most spatially and textually relevant objects. Spatio-text inverter method improves space and query efficiency. It is scalable to find the required objects. Multiple query processing method has an ability to perform more number of queries at the same time. Query processing is a set of activities involved in getting the result of a query expressed in a high level language. The query transformation should achieve both correctness and efficiency. The activities involved in retrieving data from the database.

### II. RELATED WORK

Nearest neighbor search (NNS), also known as closest point search, similarity search. It is an optimization problem for finding closest (or most similar) points. Nearest neighbor search which returns the nearest neighbor of a query point in a set of points, is an important and widely studied problem in many fields, and it has wide range of applications. We can search closest point by giving keywords as input; it can be spatial or textual.

#### A. Keyword search on spatial databases:

It focuses on finding top-k Nearest Neighbors, each node has to match the whole querying keywords. It match the whole query to each node, it does not consider the bulk of data objects in the spatial space. When number of queries increases then it leads to lower efficiency and speed. They present an efficient method to answer top-k spatial keyword queries. This work has the following contributions: 1) the problem of top-k spatial keyword search is defined. 2) The IR2-Tree is proposed as an efficient indexing structure to store spatial and textual information for a set of objects. There are efficient algorithms are used to maintain the IR2-tree, that is, insert and delete objects. 3) An efficient incremental algorithm is presented to answer top-k spatial keyword queries using the IR2-Tree.

#### B. Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems:

Location based information mostly contained in GIS database. This information contains both spatial and textual descriptions. This paper proposes an outline for GIR system and mainly focus on indexing that can process both spatial keyword queries. The following contributions in this paper: 1) It gives framework for query processing in Geo- graphic Information Retrieval (GIR) Systems. 2) Develop a novel indexing structure called KR\*-tree that captures the joint distribution of keywords in space and significantly improves performance over existing index structures. 3) This method has conducted experiments on real GIS datasets showing the effectiveness of our techniques compared to the existing solutions. It introduces two index structures to store spatial and textual information such as separate index and hybrid index.

### III. PROPOSED SYSTEM IMPLEMENTATION

The CSV (Comma Separated Values) files have been pre-processed and then it can be converted into text file then it will be stored into an data base. User can send a multiple queries (eg.search multiple dish) to the server. The spatio-text

inverter method is primarily designed to support keyword searches from a set of text files. Inverted indexes is to search file object related to specific keyword from the data base. The KcR (Keyword Count R-tree) concept is useful to find the ranking and nearest neighbour. It calculates hotel ads based on ranking and then it is to find the nearest neighbour hotel based on kilometre by using ranking hotel details. Now obtain the top-k result.

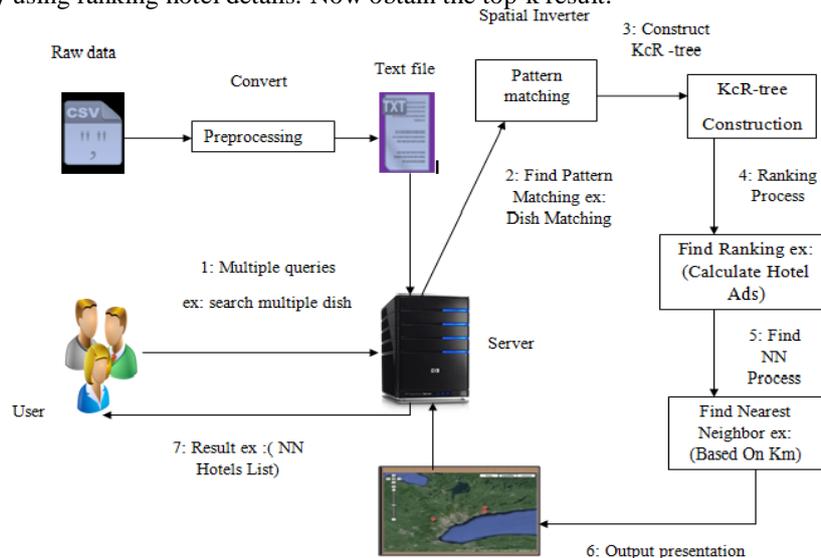


Fig. 1. Proposed System Architecture.

**A. Dataset (Preprocessing)**

The Data set preprocessing method is based on the standard format. This change CSV files to txt file. The CSV file contain raw data .It can't be a reasonable format type. Ex: the file contain word is like repeated words and special character and not orderable format so it can be changed as correct orderable format and human reading type. Then it will be load to database or server.

**B. Multiple Query Processing Methods**

The Multiple queries are very helpful to user side because they give more related details and get accurately by using top-k search methods. It will more adapt to hybrid KcR-tree format. It has the ability to perform keyword augmented in few seconds. There are two processes as ranking and nearest neighbor. It will be most facilitate to geospatial format .Next using best method is spatial inverter. It will help to multiple query concepts. It will check accurate query matching process and it give related details.

**C. Spatio-Text Inverter Method**

In Spatio -Text Inverter method, to find the matching of the query .Pattern matching is the process of checking a perceived sequence of string for the presence of the constituents of some pattern.Its efficiency derives to find a match between the search string and the text it is searching. Then, output of matching query passes to construct the KcR- tree construction.

**D. Ranking**

The KcR-tree concept is useful to find Ranking and Nearest Neighbor. It can be using Spatio-Text Inverter results. The result contain user query based hotel details .The ranking is based on hotel Ads concept and it is most helpful to user satisfaction .Then to find nearest neighbor hotel based on Kilometer by using ranking hotel details. Now obtain top-k result.

**IV. PERFORMANCE EVALUATION**

In this section, we evaluate the performance of our proposed algorithm through simulation experiments. The query locations are randomly selected in the space and the target object is selected as the t- th nearest neighbor object from the query point. For performance metrics, we measure the CPU time and the I/O cost (in terms of page accesses) in each experiment.

Table I. Cpu Time and I/O Cost Evaluation

	Existing System	Proposed System
CPU time	81%	97%
I/O cost	84%	99%

**V. CONCLUSION**

The goal of the system is to find the nearest neighbor distance based on ranking. It introduces a method to store the spatial and textual information of objects. It has a drop based technique to avoid unlimited number of query points. The

existing work isn't efficient for real time answers. The system proposes spatio-text inverter method and multiple query processing method. It has a capacity to executing more number of keyword at a time and aims to promote the ranking with least query variation. Ranking is based on hotel ad and searching user query. Both are based on counting process. Then it can be extended to customer review concept and to eliminate the fake review in future work.

#### **REFERENCES**

- [1] I. De Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In ICDE, pp. 656–665, (2008).
- [2] 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).
- [3] He.Z and Lo.E. (2012) Answering why-not questions on top-k queries. In ICDE, pp. 750-761.
- [4] Huang.W, Li.G, Tan.K and Feng.J. (2012) Efficient safe-region construction for moving top-k spatial keyword queries. In CIKM.
- [5] Korn.F and Muthukrishnan.S. (2000) Influence sets based on reverse nearest neighbor queries. In SIGMOD, pp. 201-212.
- [6] Khodaei.C, Shahabi.C, and Li.C (2010). Hybrid indexing and seamless ranking of spatial and textual features of web documents. In DEXA, pp. 450-466.
- [7] Lian.X and Chen.L. (2009). Efficient processing of probabilistic reverse nearest neighbor queries over uncertain data. VLDB J., 18(3):787-808.