

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Keyword Search for Quick Finding of Nearest Neighbor

Komal S. Bole¹, Vijay B. Gadicha², M. U. Karande³

P.G. Student, Department of CSE, VBKCOE, Malkapur, Maharashtra, India¹

Assistant Professor, Department of CSE, P. R. Patil College of Engineering & Management, Amravati,
Maharashtra, India²

Assistant Professor, Department of CSE, VBKCOE, Malkapur, Maharashtra, India³

ABSTRACT: Many applications require finding the nearest objects closest to a specified location that contains keywords. Nowadays there are lots of applications for users that purpose is to find the objects with special word or predicates & predicates with their associated text. For example, instead of considering all the hotels, a keyword search would instead searches the keywords and shows the nearest hotels that having those keywords. Existing solution to queries is based on IR2-Tree (Information Retrieval R-Tree), but it has a few shortcomings that it needs more time to process the query and fails to give real time answers. In nearest neighbor search algorithm we have used an access method called the spatial inverted index (SI Index) and SI Index improves query processing.

KEYWORDS: Keyword search, IR2- tree, spatial inverted index, nearest neighbor search.

I. INTRODUCTION

Keyword search is most popular information discovery method. Keywords are chosen words that users search. Keywords help to retrieve information faster. Indexing provides fast keyword searching. Keyword searching plays a significant role in search engine. When a set of query keywords is given by the user, the search engine searches the keyword first and returns nearer locations that containing those query keywords. For example, if user want only those hotels that containing famous dish. Means if user wants “Manchurian” only then user can enter “Manchurian” as keyword then it will return nearest hotels which has Manchurian menu.

Nearest neighbor search is an optimization problem for obtaining most similar points. Nearest neighbor is an important and broadly studied problem in various fields, and it has broad range of applications. Searching plays a significant role in searching field. Keywords along with features are taken into consideration. This approach helps to improve speed performance of information retrieval and also save searching time.

Today, the widespread use of search engines has made it realistic to write spatial queries in a brand new way. Conventionally, queries focus on objects’ geometric properties only, for instance whether a point is in a rectangle, or how close two points are from each other. Some modern applications that call for the ability to choose objects based on both of their geometric coordinates and their associated texts. For example, it would be useful if a search engine can be used to find the nearest hotels that offer “Chinese, Paneer, Manchurian” all at the same time. Note that this is not the globally nearest, but the nearest hotels among only those providing all the demanded foods.

There are simple ways to support queries that combine spatial and text features. For example, for the above query, we could first select the area in which area user has to search keyword. Firstly Search keyword after searching display the nearest hotel with distance from the user enter area to the hotel and also show time that shows how much time require to reach the hotel. This study gives importance to spatial queries with keywords [15] [16] [6]. Earlier study of

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

keyword search in relational databases is gaining importance [4] [5] [6] [10] [18]. Now this attention is diverted to multidimensional data [14] [7] [2] [8]. Some approaches use a linear ranking function [9] [8] [17] to combine spatial proximity and textual relevance.

II. LITERATURE SURVEY

It is observed that in the existing system, real nearest neighbor lies quite far away from the query point, while all the closer neighbors are missing at least one of the query keywords. IR₂-tree has a few shortcomings that affect its efficiency. The most serious one of all is that the number of false hits can be really large when the object of the final result is far away from the query point.

IR₂-TREE

The IR₂-Tree is the combination of both R tree and Signature Files. The R tree wants the more no. of keywords to search the user specification. The signature files are loading the more no text to matching the object for user specification. Here we discuss the drawback of IR₂ trees where it has the advantages of both R trees and signature files. The IR₂ trees does not contains the all the query keywords. It will direct the search to some objects those does not contain all keywords.

Signature file in general refers to a hashing-based framework, whose instantiation in is known as superimposed coding (SC). It is designed to perform membership tests: determine whether a query word *w* exists in a set *W* of words. SC is conservative, in the sense that if it says “no”, then *w* is definitely not in *W*. If, on the other hand, SC returns “yes”, the true answer can be either way, in which case the whole *W* must be scanned to avoid a false hit.

<i>word</i>	<i>hashed bit string</i>
<i>a</i>	00101
<i>b</i>	01001
<i>c</i>	00011
<i>d</i>	00110
<i>e</i>	10010

Figure 2.1 Example of bit string computation

For example, in the bit string *h(a)* of *a*, the 3rd and 5th (counting from left) bits are set to 1. As mentioned earlier, the bit signature of a set *W* of words simply ORs the bit strings of all the members of *W*. For instance, the signature of a set {*a*, *b*} equals 01101, while that of {*b*, *d*} equals 01111. Given a query keyword *w*, SC performs the membership test in *W* by checking whether all the 1's of *h(w)* appear at the same positions in the signature of *W*. If not, it is guaranteed that *w* cannot belong to *W*. Otherwise, the test cannot be resolved using only the signature, and a scan of *W* follows. A false hit occurs if the scan reveals that *W* actually does not contain *w*.

The Work done by different authors are as below:

In [6] authors focus on searching top-k nearest neighbor query. They present an effective method to answer top-k spatial keyword queries, it introduce an indexing structure called IR₂-Tree (Information Retrieval R-Tree). IR₂-tree is the most related work which is comparable to our work. Here incremental algorithm is used to reply Top-k spatial keyword queries using the IR₂-Tree. In [5] authors describe BANKS, a system which allows keyword-based search on relational databases, collectively with data and schema browsing. BANKS allows users to obtain information in an easy manner without any knowledge of the schema or any need for writing complex queries. A user can obtain information by typing a few keywords, following hyperlinks, and interacting with controls on the displayed results.

In [15] authors proposed a variant of inverted index that optimizes multidimensional points called the spatial inverted index (SI-index). In [10] authors discuss Dbexplorer, a system that enables keyword-based searches in relational databases. Dbexplorer has been implemented using a profitable relational database and Web server and allows

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

users to interact via a browser front-end. They outline the challenges and discuss the execution of our system, including results of extensive experimental evaluation.

III. PROPOSED SYSTEM

To overcome the drawbacks, a new system proposed for users. This system dealing with keyword searching and show nearer location that contains the query keyword. This system add area i.e. current location of the users so it searches the query keywords only that particular area and it gives accurate results only that area with less time. System considers domain is hotels in pune and the keywords are the menus in hotels. This system is useful for users because it trace the path from that area to the hotel on google map and also show the distance and time to reach the hotel from that area. This system is not searching the menu in hotel in the entire city, it searches only with the user enter area so it save time. System shows the result based on user location i.e. (area) and the user supplied keywords. Nearest neighbor search algorithm use an access method called the spatial inverted index (SI Index). SI-index is fairly space economical, and also it has the capability to execute keyword-improved nearest neighbor search in time.

A. ADVANTAGES OF PROPOSED SYSTEM

1. It is straight forward to enlarge our compression scheme to any dimensional space.
2. It gives result within less time so, it saves time.

B. MATHEMATICAL REPRESENTATION

Let S is the Whole System Consist of

$S = \{I, P, O\}$. Where,

I = input

$I = \{U, Q, D\}$

U is the set user.

$U = \{u_1, u_2, \dots, u_n\}$

Q is the query entered by user.

K is set of keyword entered by user.

$K = \{k_1, k_2, k_3, \dots, k_n\}$.

D is the Dataset.

NN = Nearest Neighbor

$NN = \{SI\text{-index}\}$

P = PROCESS

$P = \{NN\}$

O=Output

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

C. FLOWCHART OF PROPOSED SYSTEM

The flow of proposed system is shown in the flowchart.

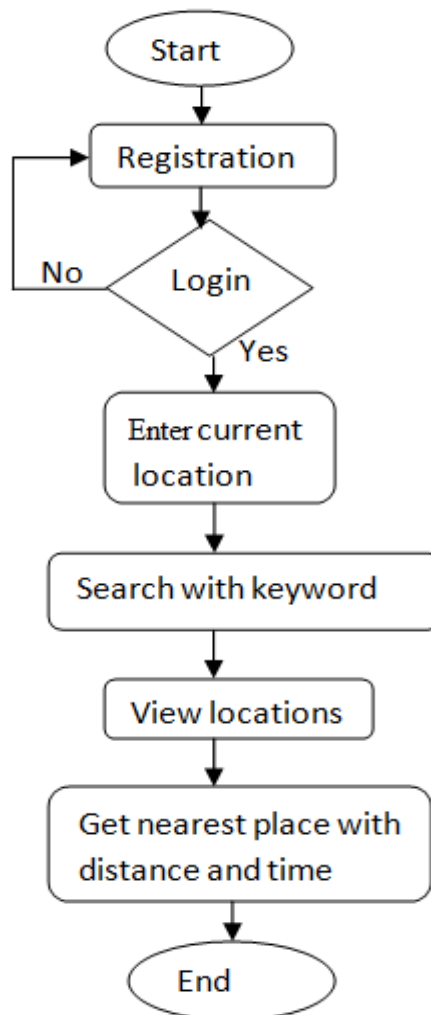


Fig 3.1: Flowchart of proposed system

D. NEAREST NEIGHBOR SEARCH ALGORITHM

- Step1: Get the enter current location (i.e. Area)
- Step 2: Search the current location
- Step 3: Read two values of current location as (Lat, Lon)
 - Where, Lat-latitude of place
 - Lon-longitude of place
- Step 4: Apply SI-index method
- Step 5: Query processing on SI-index with merging
- Step 6: Sorts all the result points
- Step 7: Get the nearest place along with distance and time having keywords.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

E. SI-INDEX (SPATIAL INVERTED INDEX) METHOD

The proposed system used the spatial inverted index method and do query processing with merging. SI-index is a compressed version of I-index with compressed coordinates. Compression eliminates the defect of I-index such that SI-index consumes much less space. Compression methods are used in order to reduce size of index where each inverted index contains only IDs. To build SI-index we have used an effective method to record gaps between consecutive IDs when compared with precise ids. For example, consider a set of integers {2, 3, 6, 8}, the gap-keeping method stores values such as {2, 1, 3, 2} illustrate in [15]. To calculate gap between ids and coordinates of points, first it takes binary form of id and coordinates of points. Then merge binary numbers to get converted values of points based on Z-curve.

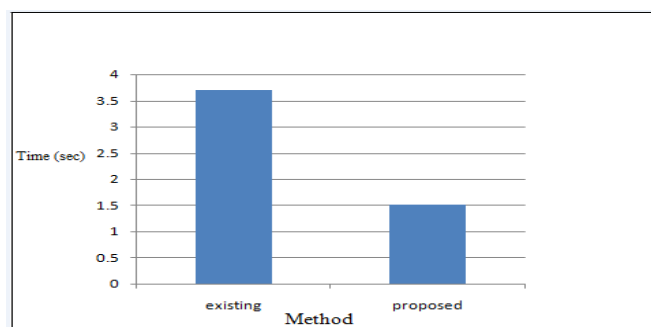
IV. RESULT

The existing and proposed methods have compared and show their graphical representation.

Table 4.1: Time for methods

Method	Time (sec)
Existing	3.7
Proposed	1.5

Table 4.1 contains the execution time of existing and proposed methods.



Graph4.1: System in terms of time

This result graph shown in the above graph4.1 indicates the performance comparison of existing and proposed methods in terms of execution time.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

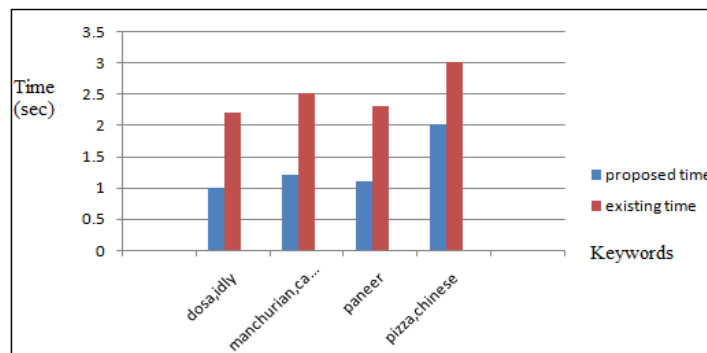
Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Table 4.2: keyword data set

Keywords	Dosa, idly	Manchurian, cake	Paneer	Pizza, Chinese
Proposed (NNS)time(sec)	1	1.2	1.1	2
Existing (IR2)time(sec)	2.2	2.5	2.3	3

Table 4.2 contains the time required to search with keywords for proposed and existing system.



Graph4.2: Query time versus number of keywords.

This result graph shown in the above graph4.2 indicates the comparison of time in seconds with different methods using additional keywords. In the above graph the keywords i.e. menu in hotel are dosa, idly, Manchurian, cake, paneer, pizza and Chinese.

V. CONCLUSION AND FUTURE WORK

This paper focuses on development of a system that searches the keywords and shows the nearest neighbor that having those keywords. The core part of this system is searching a keyword with significant information with minimum time and with suitable results. In this nearest neighbor search algorithm we have use an access method called the spatial inverted index (SI Index). SI-index is fairly space economical, and also it has the capability to execute keyword-improved nearest neighbor search in time that is at the order of dozens of milli-seconds. This method is very helpful to search the nearest neighbors that contain the keyword in real time. Searching mainly focuses on keywords. This also shows how much time taken to cover the distance of the nearest neighbor from the specific area. This application is time saving as well as it is useful to show the appropriate results to particular search.

For future work, join processing of queries can be taken into consideration algorithms.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

REFERENCES

- [1] Anusha Rachel Mammen, Elisabeth Thomas “Keyword Augmented Nearest Neighbor Search” International Journal of Engineering Research & Technology Vol.4 Issue 06, pp. 1063-1066, June-2015.
- [2] C. Usha Rani, N. Munisankar “Spatial Index Keyword Search in Multidimensional Database” International Journal of Computer Science and Information Technologies, Vol.5 (5), pp. 6468-6471, 2014.
- [3] Chandrashekhar “Fast Searching With Keywords Using Data Mining” International Journal of Computer Science and Information Technology Research, Vol.2 Issue.2, pp.82-99, April-June 2014.
- [4] Eric Chu, Akanksha Baid, Xiaoyong Chai, AnHai Doan, Jeffrey Naughton “Combining Keyword Search and Forms for Ad Hoc Querying of Databases” ACM, 2009.
- [5] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti “Keyword Searching and Browsing in Databases Using Banks” International Conference on Data engineering, 2002.
- [6] Ian De Felipe, Vagelis Hristidis, Naphtali Rische “Keyword search on spatial Databases” Proc. Int’l Conf. Data Eng. (ICDE), 2008.
- [7] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger " The R*-tree: An efficient and robust access method for points and rectangles". In Proc. of ACM, 1990.
- [8] Pranati Waghodekar, Prof. Kavita Bhosle “Survey of Efficient and Fast Nearest Neighbor Search For Spatial Query on Multidimensional Data” International Journal of Engineering Studies and Technical Approach, Vol.1 No.12, Dec-2015.
- [9] Rajkumar. R, Manimekalai. P, Mohanapriya. M, Vimalarani.C “Efficient Nearest and Score Based Ranking for Keyword Search” International Journal of Advanced Research in Computer Engineering & Technology, Volume 3 Issue 3, March 2014.
- [10] S. Agrawal, S. Chaudhari, and G. Das “Dbxplorer: A System for Keyword-Based Search over Relational Databases” International Conference on Data Engineering, 2002.
- [11] Shilpa B. patil “Survey of Searching Nearest Neighbor Based on Keywords using Spatial Inverted Index” International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014.
- [12] Sonal S. Kasare, Anup Bongale “Efficiently Searching Nearest Neighbor in Documents using Keywords”, International Journal of Engineering Research & Technology, Vol.2 Issue 12, Dec- 2013.
- [13] X. Cao, L.Chen1, G. Cong1, C. S. Jensen, Q. Qu, A. Skovsgaard, Dingming Wu, and Man Lung Yiu “Spatial Keyword Querying” Springer-Verlag Berlin Heidelberg, pp. 16-29, 2012.
- [14] X. Cao, G. Cong, and C.S. Jensen “Retrieving Top-k Prestige- Based Relevant Spatial Web Objects” International Conference on VLDB, Vol. 3, No. 1, Sep. 2010.
- [15] Y. Tao and C.Sheng “Fast Nearest Neighbor Search with Keywords” IEEE transactions on knowledge and data engineering, vol.26, no.4, April 2014.
- [16] Y.-Y Chen, T. Suel and A. Markowetz “Efficient Query Processing in Geographic Web Search Engines” ACM, June 2006.
- [17] Yufie Tao, Dimitris Papadias, Qiongmao Shen “Continuous Nearest Neighbor” VLDB Conference, 2002.
- [18] Zarreen Anjum, Priti Saktel “Survey on Nearest Neighbor Search for Spatial Database” IJCSIT Vol. 5 (6), pp.7101-7103, 2014.