

Single and Multiple point Spatial Queries Supporting Keywords for Searching Nearest Neighbors

Komal K. Chhajed

ME Student of Department of Computer Engineering MMCOE, Pune

Shailaja Jadhav

Department of Computer Engineering MMCOE, Pune

ABSTRACT

Today's applications requesting for finding spatial objects closest to a specified location or within some range which satisfy constraint of keywords. Initially, spatial queries finding nearest neighbor or range queries having conditions on only geometric properties of object points. For example, in the emergency of accident taking all hospital in consideration is not useful rather than finding hospital having facilities like ICU and emergency facilities at the same time. Currently, solution to these queries is based on the IR²-tree is not capable to provide real time effective answers. A new method named the SI-index (Spatial Inverted) is developed extends the capabilities of conventional inverted index manages with multidimensional data, along with the solution of moving range queries answered by using SI-index results to algorithm which solves the problem in real time.

Keywords

Nearest Neighbor Search, Keyword Search, Spatial Inverted Index, Range query.

1. INTRODUCTION

Today with the advent of technologies in field of Geographic Information Systems (GIS) and emerging web applications, a huge amount of geospatial data available over the web. Number of new emerging GIS applications that crawl or wrap various web sources and gives an integrated solution to the users, such as Information Mediators. A Spatial database contains points and rectangles represent multidimensional objects helps in finding spatial objects based on some specific criteria. The importance of spatial databases is that it maps the real entities in a geometric manner. As an example, location of hotels, restaurants, hospitals represented by points while combination of rectangles shows large entities like parks, lakes, and landscapes. Different functionalities of spatial database can be useful in different context. For example, in GIS, deployment of range queries as find all shopping malls in certain in which the user is moving, while nearest neighbor retrieval can discover the shopping malls closest to a given address.

Many applications require the different execution of nearest neighbor (NN) queries constrained by some properties or some adjectives like keywords of the spatial objects. On the internet providing search with keywords, today many applications provides the facility of keywords which shows the additional description to the spatial objects which satisfy the keywords constraint, in their description or other attribute. For example, if user wants to find hospitals through online yellow pages which takes address and a set of keywords as an input and returns answer in their order of distances and contains keywords in their description. Spatial keyword querying is also receiving increasing interest in the research where a range of techniques have been proposed for efficiently processing spatial keyword queries.

Spatial queries with keywords have not been extensively explored. Until this researchers are enthusiastic studying keyword search in relational databases. Now attention is diverted towards multidimensional data. Currently the solution is given by [2] which is combination of R-tree, a popular spatial index, and signature file, proved as effective method for keyword-based document retrieval. By this a structure is developed named as the IR²-tree [2], having the capabilities of both R-trees and signature files. Like R-trees, the IR²-tree preserves objects spatial proximity, a key to solving spatial queries efficiently. Another side, the IR² tree is able to alter a considerable portion of the objects that do not contain all the query keywords like signature files, which significantly reduce the number of objects to be examined.

We provide a variant of inverted index provides an optimal solution for multidimensional points called as SI (Spatial Inverted) index mechanism. The new method integrates the traditional inverted indices with point coordinates values results in compact storage facility. In the meantime, the spatial locality of data points is preserved, and results to an R-tree built over each inverted list at very less extra space overhead. As a consequence, two competing ways for query processing are provided. First is sequential merging, like conventional inverted lists are merged on the basis of ids, multiple list are merged. And second is, so leverage the R-trees, in all relevant lists browsing the points in their distances ascending order related to the query point.

The range query is one-time query that is registered with the system and is later deregistered. As long as the query remains registered, the system must send a new result to the user each time the result changes, until the solution utilizes the safe zones. When the query is computed it finds the safe zones and gives the result up to the query is registered and now we provide a solution by using conventional inverted index method used a two-step technique first compression and building R-trees which will give the answers in real time.

1.1 Motivation

With invent of new technology and emerging web with spatial or geographical objects, which are associated with user's location alongwith some textual content to the specific search. Few of these has local intent and target which are so called as spatial web objects or places, i.e. the points of interest with a web presence which have locations as well as textual contents.

A SK query i.e. spatial query alongwith the keywords takes a user location through the GPS and set of user-supplied keywords as arguments and returns objects that are simultaneously spatially and textually relevant to these arguments. With the rich semantics of geographical space and the importance of geographical space in human's life, various designs of relevant spatial keyword query functionality may be envisioned. We motivated to improve the efficiency of query processing in geographic search engines, e.g., how to maximize the query throughput for a given problem size and amount of hardware. Query processing is the major

performance bottleneck in current standard web search engines, and the main reason behind the thousands of machines used by larger commercial players and adding geographic constraints to search queries results in additional challenges during query execution.

2. RELATED WORK

Yufei et al developed method spatial inverted index that extending the inverted index to cope up with multidimensional data, and proposed algorithm which can answer the spatial queries supported with keywords in real time which can be easily incorporated with any search engine to achieve the parallelism [1]. Felipe et al proposed an efficient method which will answer top-k spatial keyword queries. A new indexing structure IR²-Tree (Information Retrieval R-Tree) is proposed by him which is combination of an R-Tree alongwith the superimposed text signatures and provides algorithms that construct and maintain an IR²-Tree, and use it to answer a top-k spatial keyword query can answer fast and provide scalability over the existing methods[2]. Ramaswamy Hariharan et al proposed a framework for GIR (Geographic Information Retrieval) systems, KR*-tree, a new data structure for processing SK (Spatial Keyword) queries alongwith the AND semantics by combined distribution of keywords appearing in space. This approach significantly reduces the drowning features which reduce the disk IOs incurred during spatial altering of objects [3]. Xin Chao et al finds the solution of retrieving a group of spatial web objects such that the group keywords matches the query keywords and such that objects are nearest to the query location and have the lowest inter object distances. He suggested two approximation algorithms with provable approximation bounds and also the exact algorithms proves to be efficient than existing methods [4]. Jensen et al proposes a new indexing framework for location aware top-k text retrieval. The proposed framework integrated inverted file and R-trees where for text retrieval inverted indices are user while R-tree are used for spatial proximity querying which proves to be efficient than existing methods. The suggested algorithms utilized the proposed indices for finding the solution of the top-k query, and takes account both text relevancy and location proximity to limit the search space at query time [5]. Yen-Yu Chen et al finds the solution for efficient query processing in scalable geographic search engines and designed a framework which rank the search results combines textual and spatial criteria, and algorithms like k-sweep algorithm, tile index algorithm, and space filling inverted index algorithm are proposed for efficiently executing ranked queries on very large collections and integrated with high-performance search engine query processor [6]. Apostolos N. Papadopoulos et al efficiently solved the problem of range queries in 2-d R-trees. It is solved by first sorting queries by means of a space filling curve, and then grouping them together and finally passed them for processing. The preprocessing of the queries lead to substantial reduction of the number of disk accesses and better cache utilization, compared to the FCFS method [7]. Lisi Chen et al gives the 12 different geo-textual indices for spatial keyword queries and provides a benchmark for indices comparison for three different types of queries as Boolean kNN queries, Top-k kNN queries, and Boolean range queries. The existing indices can be categorized according to the spatial index they utilize as R-tree based indices, grid based indices and space filling curve based indices. Using the text index employed, these indices can also be classified as inverted le based and signature le based indices. And combination of spatial and textual indices can be in category of hybrid indices [8]. Wan

D. Bae et al solves the problem of retrieving geospatial data from the web for a given spatial range query. It proposed a set of range query algorithms to completely cover the rectangular shape of the query range (completeness) while minimizing the number of k-NN searches as possible (efficiency). Based on the classification of k-NN interfaces algorithms are proposed to solve the set of range queries when the location is exposed are: 1. QDD Quad Drill Down 2. DCDT- Dynamic Constrained Delaunay Triangulation [9].

3. PROBLEM STATEMENT

To design a system to answer the spatial queries with the provision of keywords to give nearest neighbours satisfy descriptions contain the set of query keywords from a single query point and moving range queries with the exposure of user location.

Mathematically,

Let S be a dataset where every object $t \in S$ is a pair (λ', ψ') of a spatial location $t.\lambda'$ and a textual description i.e. a set of keywords $t.\psi'$.

A spatial keyword query $K = \langle \lambda', \psi' \rangle$ has two components, where $K.\lambda'$, a spatial location and $K.\psi'$, the keywords set respectively. The answer to query K is a list of k objects that are in their ascending order of distances to the query location $K.\lambda'$ and their descriptions satisfy the provided set of query keywords $K.\psi'$.

In spatial range query $K = \langle \lambda', \psi' \rangle$ has two components, where $K.\lambda'$, a spatial location and $K.\psi'$ the set of keywords respectively. The answer to query K is a list of k objects that are in ascending order of their distance to the query location $K.\lambda'$ that changes continuously and is intended to capture the user's location. The system must now send the user a new result every time the result changes due to changes in λ' and their descriptions satisfy the provided set of query keywords $K.\psi'$.

4. PROPOSED SYSTEM

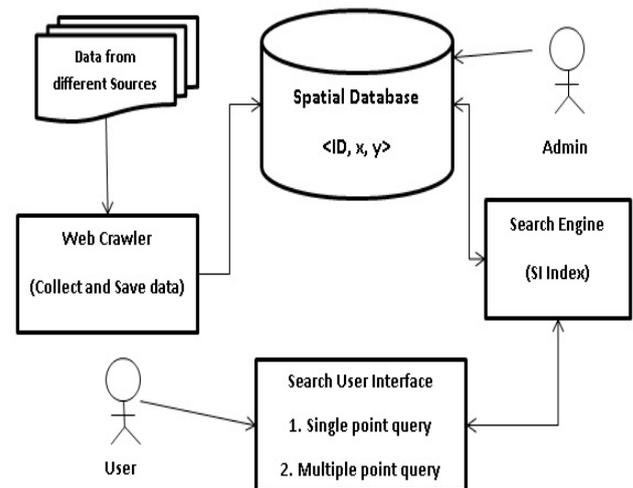


Fig1: System Architecture

The proposed GIS System is shown in Fig. 1 is divided into following modules:

1. Web crawler: It is the utility devise to collect up-to-date data from various sources over the web.

2. Spatial Database maintaining the updated information about locations points with its id and (x,y) coordinates values with all the set of keywords
3. Search Engine for fetching data from spatial database with the help of spatial inverted index method
4. Search User Interface for querying single point query and multipoint range query.

The system will as in sequence as follows and shown in fig 2:

1. Initially, the crawler fetches the data of different locations and collects the updated information from the web.
2. Then collected data is stored in spatial database maintained by admin. The information is stored with its id i.e. each location point is assigned a specific id along with its co-ordinates values i.e. x-coordinate and y-coordinate. Also the set of keywords related to point is also stored in database.
3. When user request a query from a stationary point or moving range query, the user location is taken with the help of GPS (Global Positioning System) which is processed by search engine.
4. The search engine search for nearest neighbor location from the user location which the user is requesting for, where it retrieves information with the help of SI-index (Spatial Inverted) from the database.
5. With the help of coordinates information and id values by using spatial inverted index will give the nearest neighbors which will satisfy the keywords which the user is requested.

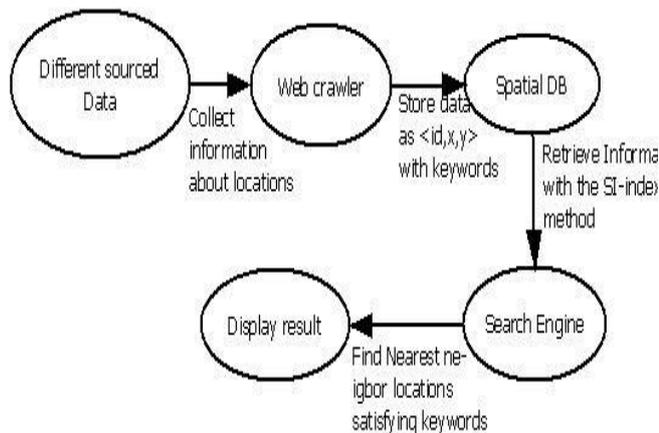


Fig 2: Data Flow Diagram of GIS

5. CONCLUSION

Many applications calls for a search engine or a search mechanism which can find spatial queries supported with keywords. With the increase in use of GPS enabled devices, users are interested in finding required locations within least amount of time. The existing solutions are space consumptive so which are unable to give real time answers. They were unable to give specific location within expected amount of time and were more space consumptive. So for this, a solution called as Spatial-Inverted Index (SI-Index) is provided which is an indexing technique which can answer the queries from the single query point. Also alongwith this, solution for moving range queries with keywords is provided by safe region assignment algorithm alongwith the spatial inverted

index technique which reduced the computation cost and provide the answers in very few milli-seconds. For future work we can incorporate our search to solve queries like collective spatial keyword queries, queries with different degree of continuity or spatial queries provided with keywords or more specific locations can be search within very few seconds of time combining different techniques.

6. ACKNOWLEDGEMENT

We take this opportunity to thank Prof. Ram Joshi for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this paper. We are also thankful to all the staff members of the Department of Computer Engineering for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

7. REFERENCES

- [1] Yufei Tao and Cheng Sheng, "Fast Nearest Neighbor Search Using Keywords", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.
- [2] Ian De Felipe Vagelis Hristidis Naphtali Rische, "Keyword search in spatial databases ", IEEE Netw. Vol. 23, no. 1, pp. 4247, Jan. 2009.
- [3] Ramaswamy Hariharan, Bijit Hore, Chen Li, Sharad Mehrotra, "Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval(GIR) Systems", IEEE Trans. Veh. Technol., vol. 55, no. 4, pp. 13851396, Jul. 2006.
- [4] Xin Cao, Gao Cong, Christian S. Jensen, Beng Chin Ooi, "Collective Spatial Keyword Querying" , IEEE Power Energy Mag, vol. 10, no. 1, pp. 5866, Feb. 2012
- [5] 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.
- [6] Yen-Yu Chen, Torsten Suel, Alexander Markowetz, "Efficient Query Processing in Geographic Web Search Engines", IEEE TRANSACTIONS ON SMART GRID, VOL. 2, NO. 4, DECEMBER 2011
- [7] Apostolos N. Papadopoulos and Yannis Manolopoulos, "Multiple Range Query Optimization in Spatial Databases", IEEE TRANSACTIONS ON SMART GRID, VOL. 2, NO. 4, DECEMBER 2011.
- [8] Lisi Chen, Gao Cong, Christian S. Jensen, Dingming Wu, "Spatial Keyword Query Processing: An Experimental Evaluation" IEC TS 62351-1, Power Systems Management and Associated Information Exchange Data and Communications Security, May 2007.
- [9] Wan D. Bae, Shayma Alkobaisi, Seon Ho Kim, Sada Narayanappa, Cyrus Shahabi, " Web data retrieval: solving spatial range queries using k-nearest neighbor searches", IEEE Spectrum vol. 50, no. 3, pp. 4853, Mar.2013.