International Journal Of

Recent Scientific Research

ISSN: 0976-3031

Volume: 7(11) November -2015

FAST NEAREST NEIGHBOR SEARCH WITH KEYWORDS FOR SPATIAL DATASET

Akshaykumar Landge, Kailas Mehtre, Atul Kadam and Niket Dodal

THE OFFICIAL PUBLICATION OF
INTERNATIONAL JOURNAL OF RECENT SCIENTIFIC RESEARCH (IJRSR)
http://www.recentscientific.com/ recentscientific@gmail.com

ISSN: 0976-3031

Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research
Vol. 6, Issue, 11, pp. 7388-7390, November, 2015

International Journal of Recent Scientific Research

RESEARCH ARTICLE

FAST NEAREST NEIGHBOR SEARCH WITH KEYWORDS FOR SPATIAL DATASET

Akshaykumar Landge*, Kailas Mehtre, Atul Kadam, Niket Dodal Kishori Pawar and Deepalee Chaudhari

Akurdi Railway station Road, Gurudwara Chowk, Sai Sakshi Hotel

ARTICLE INFO

Article History:

Received 16th August, 2015
Received in revised form 24th September, 2015
Accepted 23rd October, 2015
Published online 28st November 2015

Key words

Keyword search, Nearest neighbors, Location-based search, Spatial databases, Spatial index.

ABSTRACT

Many modern applications call for novel forms of queries that aim to find objects satisfying both a spatial predicate, and a predicate on their associated texts .Currently, the best solution to such queries is based on the IR2 -tree, which has a few deficiencies that seriously impact its efficiency. Motivated by this, we develop a new access method called the spatial inverted index that extends the conventional inverted index to cope with multidimensional data, and comes with algorithms that can answer nearest neighbor queries with keywords in real time.

INTRODUCTION

A spatial database oversees multidimensional items and gives quick access to those items in view of diverse choice criteria. The significance of spatial databases is reflected by the accommodation of demonstrating elements of reality in a geometric way. For instance, areas of eateries, inns, doctor's facilities thus on are frequently spoken to as focuses in a guide, while bigger degrees, for example, stops, lakes, and scenes regularly as a mix of rectangles. Numerous functionalities of a spatial database are helpful in different courses in particular contexts. For occurrence, in a topography data framework, range pursuit can be sent to discover all eateries in a certain region, while closest neighbor recovery can find the eatery nearest to a given location.

Today, the across the board utilization of web indexes has made it reasonable to compose spatial inquiries in a brand new manner. Ordinarily, inquiries concentrate on objects' geometric properties just, for example, whether a point is in a rectangle, or how close two focuses are from one another. We have seen some present day applications that require the capacity to choose articles in light of both of their geometric directions and their related writings. For instance, it would be genuinely valuable if a pursuit motor can be utilized to locate the closest eatery that offers "steak, spaghetti, and liquor" all in the meantime. This is not the "internationally" closest eatery (which would have been returned by a customary closest neighbor question), however the closest eatery among just those giving all the requested sustenance and beverages.

There are simple approaches to bolster questions that consolidate spatial and content elements. For instance, for the above inquiry, we could first get every one of the eateries whose menus contain the arrangement of decisive words {steak, spaghetti, brandy}, and after that from the recovered eateries, discover the closest one. Additionally, one could likewise do it conversely by focusing on first the spatial conditions peruse every one of the eateries in rising request of their separations to the question point until experiencing one whose menu has all the magic words. The significant downside of these direct methodologies is that they will neglect to give constant answers on troublesome inputs. A regular case is that the genuine closest neighbor lies entirely far away from the question point, while all the closer neighbors are lost no less than one of the inquiry magic word.

Corresponding author;* **Akshaykumar Landge
Akurdi Railway station Road, Gurudwara Chowk, Sai Sakshi Hotel

Literature survey

In [1] Internet search engines have popularized the keyword-based search paradigm. While traditional database management systems offer powerful query languages, they do not allow keyword-based search. In this paper, we discuss DBXplorer, a system that enables keyword-based searches in relational databases.

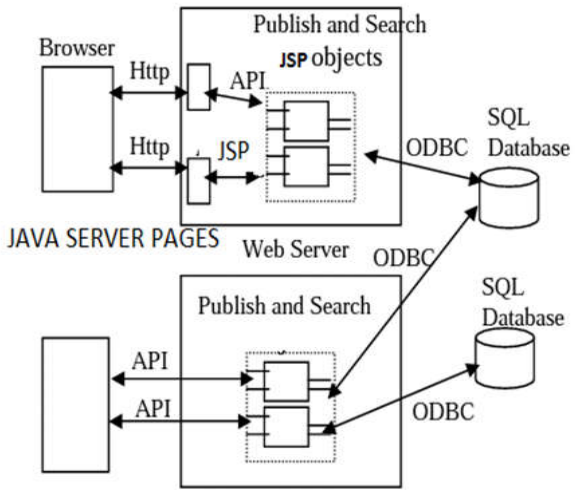

Fig.1 System Architecture

DBXplorer has been implemented using a commercial relational database and Web server and allows users to interact via a browser front-end. We outline the challenges and discuss the implementation of our system, including results of extensive experimental evaluation.

The paper [2] With the growth of the Web, there has been a rapid increase in the number of users who need to access online databases without having a detailed knowledge of the schema or of query languages; even relatively simple query languages designed for non-experts are too complicated for them. We describe BANKS, a system which enables keyword-based search on relational databases, together with data and schema browsing. BANKS enables users to extract information in a simple manner without any knowledge of the schema or any need for writing complex queries. A user can get information by typing a few keywords, following hyperlinks, and interacting with controls on the displayed results. BANKS models tuples as nodes in a graph, connected by links induced by foreign key and other relationships. Answers to a query are modeled as rooted trees connecting tuples that match individual keywords in the query. Answers are ranked using a notion of proximity coupled with a notion of prestige of nodes based on inlinks, similar to techniques developed for Web search. We present an efficient heuristic algorithm for finding and ranking query results.

In [3] The web is increasingly being used by mobile users. In addition, it is increasingly becoming possible to accurately geo-position mobile users and web content. This development gives prominence to spatial web data management. Specifically, a spatial keyword query takes a user location and user-supplied keywords as arguments and returns web objects that are spatially and textually relevant to these arguments. This paper reviews recent results by the authors that aim to achieve spatial keyword querying functionality that is easy to use, relevant to users, and can be supported efficiently. The paper covers different kinds of functionality as well as the ideas underlying their definition.

This paper [4] location-aware keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords. This query occurs inherently in many types of mobile and traditional web services and applications, e.g., Yellow Pages and Maps services. Previous work considers the potential results of such a query as being independent when ranking them. However, a relevant result object with nearby objects that are also relevant to the query is likely to be preferable over a relevant object without relevant nearby objects. Proposes the concept of prestige-based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects. Based on this, a new type of query, the Location-aware top-k Prestige-based Text retrieval (LkPT) query, is proposed that retrieves the top-k spatial web objects ranked according to both prestige-based relevance and location proximity. We propose two algorithms that compute LkPT queries. Empirical studies with real-world spatial data demonstrate that LkPT queries are more effective in retrieving web objects than a previous approach that does not consider the effects of nearby objects; and they show that the proposed algorithms are scalable and outperform a baseline approach significantly.

In [5] Geographic web search engines allow users to constrain and order search results in an intuitive manner by focusing a query on a particular geographic region. Geographic search technology, also called local search, has recently received significant interest from major search engine companies. Academic research in this area has focused primarily on techniques for extracting geographic knowledge from the web. In this paper, we study the problem of efficient query processing in scalable geographic search engines. Query processing is a major bottleneck in standard web search engines, and the main reason for the thousands of machines used by the major engines. Geographic search engine query processing is different in that it requires a combination of text and spatial data processing techniques. We propose several algorithms for efficient query processing in geographic search engines, integrate them into an existing web search query processor, and evaluate them on large sets of real data and query traces.

Existing system

Spatial queries with keywords have not been extensively explored. In the past years, the community has sparked enthusiasm in studying keyword search in relational databases. It is until recently that attention was diverted to multidimensional data. The best method to date for nearest neighbor search with keywords is due to Felipe *et al.*. They nicely integrate two well-known concepts: R-tree, a popular spatial index, and signature file, an effective method for keyword-based document retrieval. By doing so they develop a structure called the IR2 -tree, which has the strengths of both R-trees and signature files. Like R-trees, the IR2 - tree

preserves objects' spatial proximity, which is the key to solving spatial queries efficiently. On the other hand, like signature files, the IR2 -tree is able to filter a considerable portion of the objects that do not contain all the query keywords, thus significantly reducing the number of objects to be examined.

Proposed system

In this paper, we design a variant of inverted index that is optimized for multidimensional points, and is thus named the spatial inverted index (SI-index). This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme. Meanwhile, an SI-index preserves the spatial locality of data points, and comes with an R-tree built on every inverted list at little space overhead. As a result, it offers two competing ways for query processing. Merge multiple lists very much like merging traditional inverted lists by ids. Alternatively, we can also leverage the R-trees to browse the points of all relevant lists in ascending order of their distances to the query point. As demonstrated by experiments, the SI-index significantly outperforms the IR2 -tree in query efficiency, often by a factor of orders of magnitude. Distance browsing is easy with R-trees. In fact, the best-first algorithm is exactly designed to output data points in ascending order of their distances. It is straight forward to extend our compression scheme to any dimensional space.

Neighbor search

This module we implement our neighbor Search. The other problem with this search algorithm is that the indexing information has to be replicated in the broadcast cycle to enable twice scanning. The first scan is for deciding the search range, and the second scan is for retrieving k objects based on the search range. Therefore, we propose the Nearest Neighbor query approach to improve algorithm.

CONCLUSION

In this paper plenty of applications calling for a search engine that is able to efficiently support novel forms of spatial queries that are integrated with keyword search. The existing solutions to such queries either incur prohibitive space consumption or are unable to give real time answers. In this paper, we have remedied the situation by developing an access method called the spatial inverted index (SI-index). Not only that the SI-index is fairly space economical, but also it has the ability to perform keyword-augmented nearest neighbor search in time that is at the order of dozens of milli-seconds. Furthermore, as the SI-index is based on the conventional technology of inverted index, it is readily incorporable in a commercial search engine that applies massive parallelism, implying its immediate industrial merits.

Acknowledgment

We express our sincere gratitude to Mrs. M. A Potey, Head of Department, Computer Engineering, for her assistance, persuasion, and an incent to work better.

Our deepest gratitude goes to our project guide, Mrs. D. A. Chaudhari, for her guidance, ideas, help, encouragement, interpretations and suggestions which helped us in the realization of our objective and coordinate as a team. We wish to thank Mrs. Kishori Pawar for their constant guidance. We are extremely thankful to our project incharge, Mrs. Shanthi Guru for her comments, introspections and support which helped us throughout our work.

Reference

1. S. Agrawal, S. Chaudhuri, and G. Das, "Dbxplorer: A System for Keyword-Based Search over Relational Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 5-16, 2002.
2. G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S.Sudarshan, "Keyword Searching and Browsing in Databases Using Banks," Proc. Int'l Conf. Data Eng. (ICDE), pp. 431-440, 2002.
3. X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M.L. Yiu, "Spatial Keyword Querying," Proc. 31st Int'l Conf. Conceptual Modeling (ER), pp. 16-29, 2012.
4. X. Cao, G. Cong, and C.S. Jensen, "Retrieving Top-k Prestige- Based Relevant Spatial Web Objects," Proc. VLDB Endowment, vol. 3, no. 1, pp. 373-384, 2010.
5. Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processingin Geographic Web Search Engines," Proc .ACM SIGMOD Int'l Conf. Management of Data, pp. 277-288, 2006.
6. Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid Index Structures for Location-Based Web Search," Proc. Conf. Information and Knowledge Management (CIKM), pp. 155-162, 2005.
7. 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.
8. D. Conte, P. Foggia, M. Petretta, F. Tufano, and M. Vento, "Meeting the application requirements of intelligent video surveillance systems in moving object detection," in Pattern Recognition and Image Analysis, ser. LNCS. Springer Berlin Heidelberg, 2005, vol. 3687, pp. 653–662
9. J.S. Vitter, "Algorithms and Data Structures for External Memory," Foundation and Trends in Theoretical Computer Science, vol. 2, no. 4, pp. 305-474, 2006.
10. 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.

****

ISSN 0976-3031