Crawling Hidden Objects with KNN Queries

[1]Pavankumar Naik, [2]Basavaraj Muragod ,[3]Prema M ,[4]Renuka M R , [5]Aaisha R M ,[6]Malati C K

[1,2]Assistant Professor, Dept. of Computer Science & Engineering, SKSVMACET, Karnataka, India
[3,4,5,6] Student, Dept. of Computer science and Engineering, SKSVMACET, Karnataka, India

---***---

Abstract – *Through Many websites we are finding Location Based Services (LBS) that provide a top k Nearest Neighbours objects (e.g.,shortest restaurants) for a given query location. This paper deals with the information about finding the problem of crawling (searching) all objects efficiently from an LBS website, it provides the public kNN web search interface. Exactly, we develop 2D and higher-dimensional spaces by using crawling algorithm, respectively, and demonstrate that our algorithms overhead by a function of the number of dimensions and the number of crawled objects, in any case of the main distributions of the objects. We also enlarge the crawling algorithms to control the scenarios where definite backup information about the basic data distribution, e.g., the society density of an area which is often positively related with the density of LBS objects, is available.*

Key Words: Information and Communications Technology, Location based Searching technique, LBS server, kNN queries, crawling objects.

1. INTRODUCTION

Based on increasing popularity, Location Based Services (LBS), e.g., Google Maps, Yahoo Local etc., started offering web-based search technique that resemble a kNN query interface. Specifically, for a user-specified query location q, these websites fetching from the objects in its backend data the top-k nearest neighbors to q and back to these k objects to the user through the web interface. Here k is often a small value like 50 or 100. For example, Mc-Donald's [1] returns the top 25 nearest restaurants for a user-specified location through its locations search webpage. While such a kNN (crawling nearest neighbour) search interface is often sufficient for an individual user can easily get the nearest restaurants and researchers concentrate in an LBS service often desire a more comprehensive view of its underlying data.

This is an unavoidable artifact of the space-partitioning strategy taken by the two techniques - one using Quad Tree whiles the other using Constrained Delaunay Triangulation. Nonetheless, as we shall show in the experimental results, it may lead to serious efficiency problems while running the algorithms in practice, especially when the space size is large but the desired objects are few and congregated in small clusters. Another problem shared by both existing techniques is that they only work on 2D spaces, but not higher dimensional spaces that expose a KNN interface.

1.1 Hidden Data

Through many available public domains we gather external knowledge, which can effectively indicate the distributions of hidden objects (points) in the space. For example, the number of restaurants is highly related to the distribution of population, or road densities of regions. In this section, we use a 2- D kNN spatial database of restaurants as an example, the crawling algorithm helps to study and how to use road information to improve our external knowledge. Through crawling algorithms we also find the scalability, with different size of the databases from the figure. Besides, it costs more queries to crawl all points when the hidden points are in skewed distribution.

1.2 Data Crawling

Crawling algorithm with external Source: The Two-D crawling algorithm is performed after partitioning the Two-D space using external Source. This is one of the most crawling (searching) algorithms this paper proposed in Two-D space. The DCDT crawling algorithm: This algorithm was proposed in work. To our source, this crawling algorithm is the advanced for kNN based databases in 2-D space. The constrainedDelaunaytriangulation technique is implemented by authors, partitions the uncovered regions into triangles, and then the new query is taken on the Centre of the toughest triangle. Their algorithm recursively repeated this process until no uncovered triangles are left. We can find the measurability of the algorithms with different size of the databases.

1.3 Location Based Services

Location Based Services (LBS), e.g., WeChat, FourSquare, etc., started offering web-based search techniques that resemble a kNN query interface. Specifically, for a user-specified query location q, these websites fetching from the objects in its backend data to the top-k nearest neighbour to q and return back to these k objects to the user through the web interface. In this paper, we study the problem of crawling the LBS through the restricted kNN search interface. Although hidden points usually exist in 2-D space, there are some applications with points in higher dimensional spaces. We extend the 2-D crawling algorithm to the general m-D space, and give the m-D crawling algorithm with theoretical upper bound analysis. This paper addresses the problem of crawling all objects efficiently from an LBS website, through the public kNN web search interface it provides. Specifically, we develop crawling algorithm for 2D and higher-dimensional spaces, respectively, and demonstrate through theoretical analysis that the overhead of our algorithms can be bounded by a function of the number of dimensions and the number of crawled objects, regardless of the underlying distributions of the objects.

1.3 KNN Queries

Web-based search technique provides a kNN (crawling nearest neighbour) query. exactly, for a user-specific query location q, these websites extract the objects form in its backend database of the top-k nearest neighbors(KNN) to q and return back to these k objects to the user through the web based. KNN search is often help full for an individual user looking for the nearest restaurants researchers are interested in an LBS(Location based service) service often desire a more comprehensive view of its underlying data. It is important that the key technical challenge for crawling through a kNN interface is to minimize the more number of queries issued to the LBS service. The requirement is by limitations imposed by most LBS services on the number of queries deals from an IP address or a user account (in case of an API service such as Google Maps) for a given time period (e.g., one day).

2. LITERATURE SURVEY

Thomas F. La Porta, Yan Sun introduced a Location-Based Services System (LBSs) for location sharing in social networks. LBS system is used to protect the personal details of the user locations. It protects user uniqueness within basic mobile services.

This focuses on following aspects: User should be control the access to location information at different levels of granularity and with different levels of user control, user has to describe the cluster of entity that are allowed to access its location information LBS support location privacy control by the user. It supports user control and scalability. It provides Instant Messaging service for server and clients [3].

Weijia Jia and Ke introduced an unnamed confirmation protocol based on unnamed proxy for wireless systems. Roaming user does not like to identify their own information to other user; they want to protect their information while wondering from home network to foreign network [4].

Controlling personal user location under untrusted server may cause the privacy problem for the user in wireless sensor network. For this reason Chi-Yin Chow, Mohamed F. Mokbel, and Tian introduces a preserving-privacy user location controlling system to provide better security to the user. Chi-Yin Coweta proposes a two in-network algorithm, which are information and quality-aware algorithms used to protect the location information of the user [5].

An additional form of web search, known as online shortest path search, is trendy due to progress in geo-positioning. Nevertheless, existing caching approaches are unsuccessful for shortest path queries. This is because of numerous crucial differences between web search results and shortest path results, in familiar to query matching, cache point overlapping, and query cost difference. Motivated by this, they identify several properties that are essential to the success of effective Caching for shortest path search.

3. PROPOSED SYSTEM

We develop crawling algorithm for 2D and higher-dimensional spaces, respectively, and demonstrate through theoretical analysis that the overhead of our algorithms can be bounded by a function of the number of dimensions and the number of crawled objects, regardless of the underlying distributions of the objects.

- Then we develop our OPTIMAL-1D-CRAWL algorithm for databases in 1-D spaces which can avoid the above mentioned problem.
- Finally, we give the theoretical analysis of the proposed algorithm. Above theorem shows that the proposed crawling algorithm can perform with cost linearly related to the number of points of the database if the point density in the region changes not too much.
- We also checked the proposed crawling algorithms on the real data sets Yahoo Local in 2-D space and Eye-glasses in 4-D space.
- We explained the details of these datasets respectively as follows ,This algorithm is the state of the art of crawling(searching) algorithm for kNN based databases in Two-D space.
- Constrained Delaunay triangulation this technique is implemented by authors in their work to always partition the uncovered regions into triangles, then issued the new query on the center of the biggest triangle.

OPTIMAL-1D-CRAWL Algorithm:

The Algorithm 1 is having all details about OPTIMAL-1DCRAWL and this algorithm targets the midpoints of uncovered regions while the previously described overlapping algorithm targets the boundaries of uncovered regions - just this subtle difference leads to fundamentally different query complexity results.

DBSCAN for grids clustering:

This paper introduced a new algorithm GRPDBSCAN (Grid-based DBSCAN Algorithm with Referential Parameters). GRPDBSCAN, which gather the grid partition technique and multiple-density based on the clustering algorithm, improved its efficiency. On the other hand, the Eps and Minpts parameters of the DBSCAN algorithm were they auto-generated, more objective.

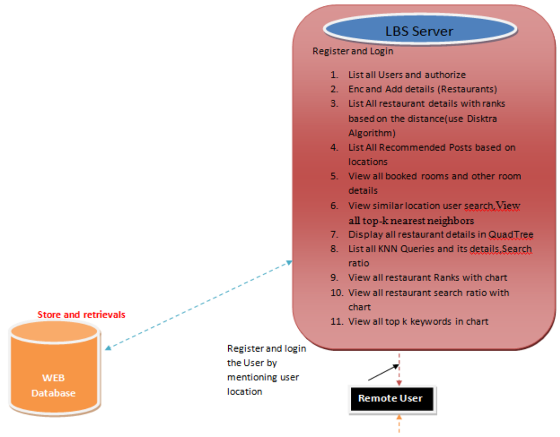
LBS Server
Register and Login
1. List all Users and authorize
2. Enc and Add details (Restaurants)
3. List All restaurant details with ranks based on the distance(use Disktra Algorithm)
4. List All Recommended Posts based on locations
5. View all booked rooms and other room details
6. View similar location user search,View all top-k nearest neighbors
7. Display all restaurant details in QuadTree
8. List all KNN Queries and its details,Search ratio
9. View all restaurant Ranks with chart
10. View all restaurant search ratio with chart
11. View all top k keywords in chart
Store and retrievals
WEB Database
Register and login the User by mentioning user location
Remote User

Figure 3.1: Architecture Diagram

Objective of the project:

Our objective in this project is to enable the crawling of an LBS database by issuing a small number of queries through its publicly available kNN(Crawling Nearest Neighbour) web search technique, so that afterwards a data searchers can simply treat the searched data as an offline database and perform whatever diagnostic operations it desired. Here "crawling" (searching)is mainly defined, i.e., it can refer to the extraction of all objects from the database ,or only those objects that satisfy certain selection conditions, so long as such conditions can be "passed through" to the kNN interface.

4. DESIGN

Design includes the initially required diagrams such as flowcharts, ER diagram to develop the system.

4.1 Flow Chart

This flow chart illustrates organization chart, or similar formalized structure. The flowchart is a means to visually present the flow of data through an information processing systems.

Admin Login:

The figure 4.1 shows in this flowchart for admin login has to login with valid username and password. After admin login successful he can do some actions such as view all user, their details and authorize them , add restaurants and its details(Encrypt the details), list all restaurants and its details using disktra algorithm, View all recommended restaurants based on locations, view all booked rooms and its details, View all top k nearest neighbors(keywords), view all restaurants in quad tree, List all query keywords and its search ratio, view all restaurant ranks in a chart, view all restaurant search ratio in chart, view all top k keywords in a chart.

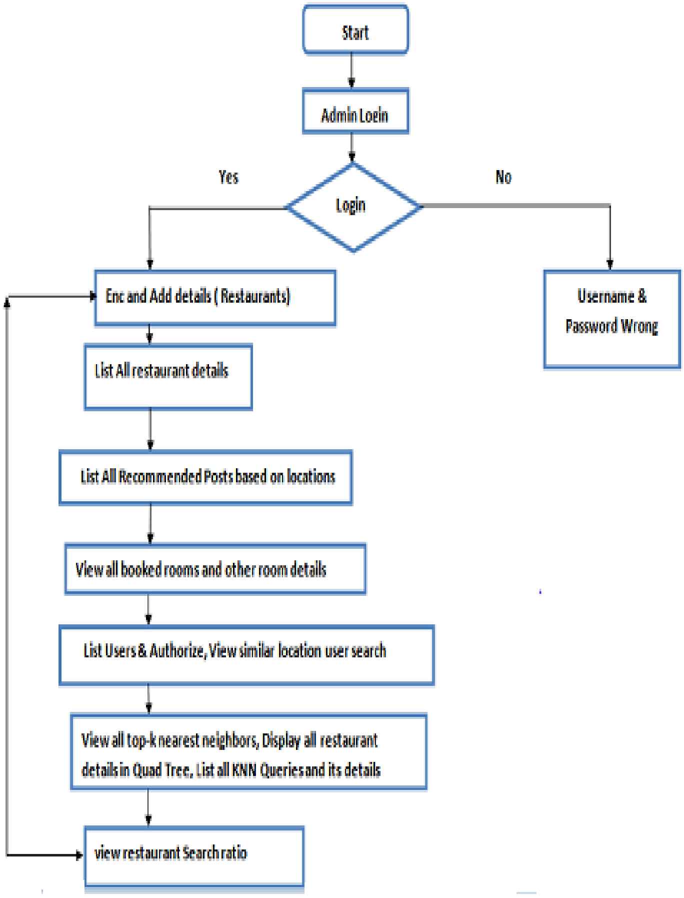
Start
Admin Login
Login
Yes
No
Enc and Add details (Restaurants)
Username & Password Wrong
List All restaurant details
List All Recommended Posts based on locations
View all booked rooms and other room details
List Users & Authorize, View similar location user search
View all top-k nearest neighbors, Display all restaurant details in Quad Tree, List all KNN Queries and its details
view restaurant Search ratio

Figure 4.1: Flowchart for Admin login

User Login:

The figure 4.2 shows the flowchart for user login. There are n numbers of users are present. User should register before doing some operations after registration successful he can entered by using valid user(customer) name and password. Login successful he will do some operations like view profile details, search restaurants based on your location (using Dijkstra algorithm), show it in Gmap using multiple markers and recommend to other users, and also display similar links which are present in other location, view crawling time (search time), book the rooms, Search restaurants based on keyword, view all recommended restaurants, view search keyword and its ratio.

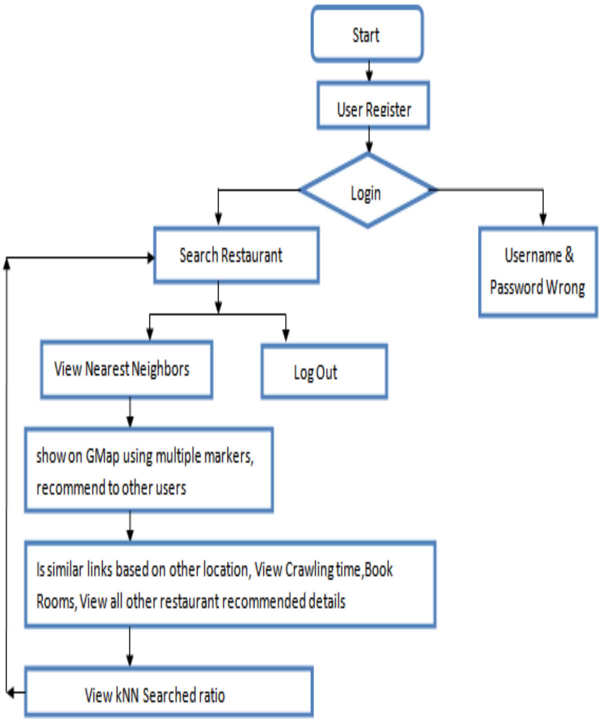
Start
User Register
Login
Search Restaurant
Username & Password Wrong
View Nearest Neighbors
Log Out
show on GMap using multiple markers, recommend to other users
Is similar links based on other location, View Crawling time,Book Rooms, View all other restaurant recommended details
View kNN Searched ratio

Figure 4.2: Flowchart for User login

4.2 ER Diagram

An entity relationship model, another meaning an entity relationship(ER) diagram is a graphical entity is a symbol of entities and their coordination to each other, naturally used in computing is consider to the organization of data within database or information systems. An entity is piece of data is an object or concept about which information is stored.ER diagram looks very much like a flowchart. It is the specialized symbols, and the meanings of those symbols, that make it unique.

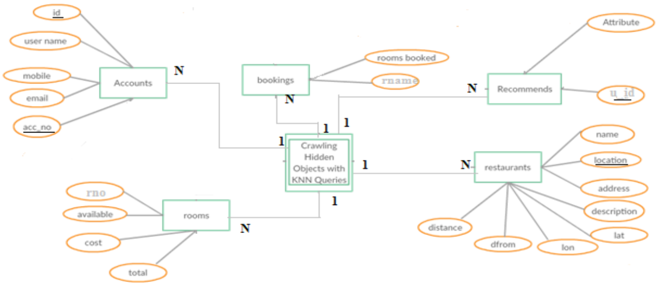

Figure 4.3: Entity Relationship Diagram

5. IMPLEMENTATION

Implementation deals with the tools used for front end design and techniques used for back end connections. Eclipse is the tool on which the web application is developed. The other tools used are MySQL and JDK. We are using the programming languages HTML and Java. The programming techniques used are Angular JS and Bootstrap.

AngularJS (commonly referred to as "Angular.js") is a JavaScript-based open-source front-end web application framework mainly maintained through Google and by a community of individuals and corporations to address many of the challenges encountered in developing single-page applications. The Angular JS framework gives more important reading the HTML page, which has embedded into it additional custom tag attributes. Bootstrap is a free and open-source front-end web framework for designing websites and web applications. It contains HTML- and CSS-based design templates for forms, buttons etc. and other interface components, as well as optional JavaScript extensions. Unlike most web frameworks, it concerns about its front-end development only.

The tables are created once in MySQL command prompt. Connection between front end to backend is done by Hibernate. Hibernate is open source Java Framework. It's primary feature is mapping from Java classes to database tables.

By using the above concepts we implemented web application through three steps:

Step 1: Development of Web Application using html, css and Java.
Step 2: Creating tables in MySQL command line prompt.
Step 3: Hosting the Application in cloud and Running in browser.

Implementation is the main stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most difficult stage in achieving a successful new system and it should be work with confidence and effective. This new system is giving to user.

The implementation stage involves careful planning, testing of the existing system and it's constraints on implementation, designing of methods to achieve changing and evaluation of changeover methods.

Modules:

In this Project we implemented in four modules

i) Hidden Data,

ii) Data Crawling,

iii) Location Based Services,

iv) KNN Queries.

6. OUTCOMES FROM PROPOSED SYSTEM

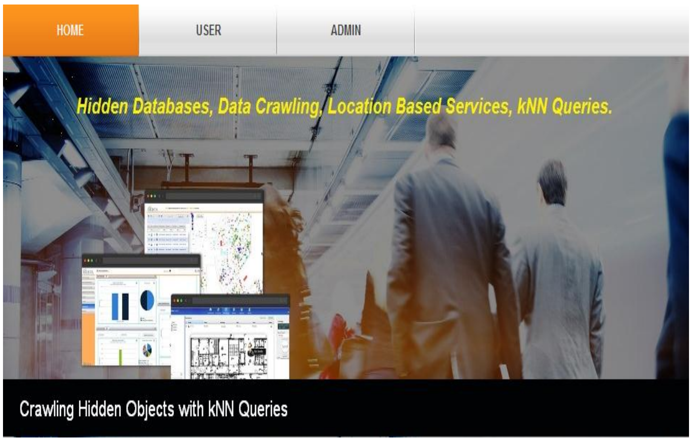
HOME
USER
ADMIN
Hidden Databases, Data Crawling, Location Based Services, kNN Queries.
Crawling Hidden Objects with kNN Queries

Figure 6.1: Login Page

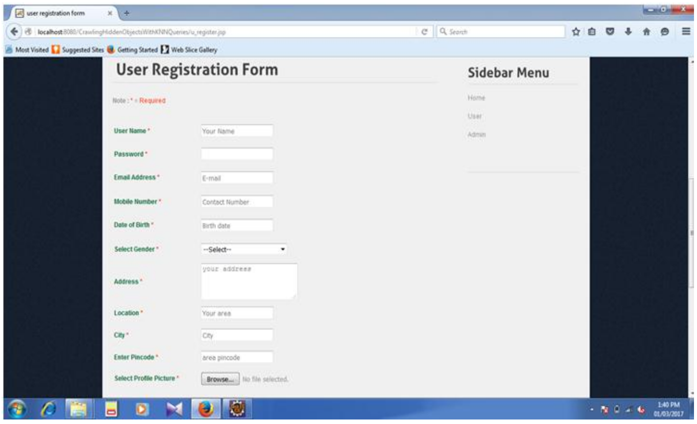

Figure 6.2: Register form

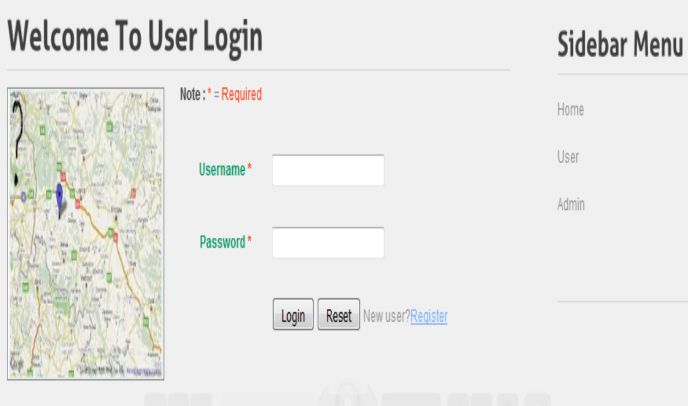

Figure 6.3: User Login

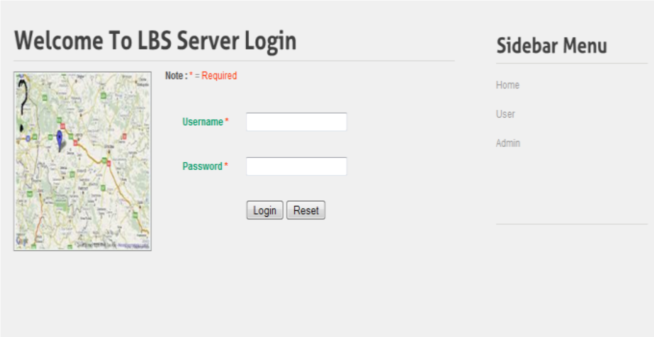

Figure 6.4: LBS Server Login (Admin Login)

6. CONCLUSION

The problem of searching the LBS(Location based services) through the restricted kNN search interface. Although hidden points usually exist in 2-D space, there are some applications with points in higher dimensional spaces. We expand the Two-D crawling algorithm to the controller of m-D space, and give the m-D crawling (searching) algorithm with theoretical higher bound analysis. For 2-D space, we take external knowledge into consideration to improve the crawling performance. The proposed system outcomes show the success of our algorithms. In this study, the proposed algorithms crawl data objects by given a rectangle (cube) in the spatial space. In this circumstance when the bounded region of the objects is irregular, it can be pre-partitioned into a set of rectangles (cubes) before using the techniques expected in this paper.

REFERENCES

[1] Yu Li and Man Lung Yiu, “Route(path)-Saver: influence Route APIs for Accurate and Efficient Query Processing at Location-Based Services”.

[2] Q. Han, Y. Zhang, X. Chen, H. Li, and J. Quan, “Efficient and robust identity-based handoff authentication in wireless mobile networks,” in Proc. 6th Int. Conf. Network and System Security, Fujian, China, 2012, pp. 180–191, LNCS 7645.

[3] H.-P. Kriegel, P. Kroger, M. Renz, and T. Schmidt, “Proximity queries in large traffic networks,” in Proc. 15th Annu. ACM Int. Symp. Adv. Geographic Inform. Syst., 2007.

[4] D. Papadias, J. Zhang and Y. Tao, “Query development in four dimensional network databases,” in Proc. 29th Int. Conf. Very huge Data Bases, 2003.

[5] H. Samet, J. Sankaranarayanan “Scalable network distance browsing in four-dimensional databases,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2008.

[6] Y. Sun, T. La Porta "A elastic privacy enhanced location-based services system framework and practice," IEEE (Institute of Electrical and Electronics of Engineering) Trans. Mobile Comput., Mar. 2009.

[7] Chunlin Jiang, Weijia JIA and Ke GU, "An Anonymous Wireless Authentication Protocol mainly concentrates on Proxy Signature". Trans. 2012.

[8] C.Y Chow, M.F.Mokbel, and T. He, "A privacy-preserving location monitoring system for wireless sensor networks," IEEE Trans. Mobile Comput., vol. 10, no. 1, Jan. 2011.

[9] Jeppe Rishede Thomsen Man Lung Yiu, "Effective Caching of Shortest Paths for Location-Based Services".

[10] Femi Olumofin, Piotr K. Tysowski, Ian Goldberg, and Urs Hen Gartner doing Efficient Query Privacy for Location Based Services.