ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 3, Issue 4, April 2018

Efficient Query Processing In Geographic Web Search Engines

[1] R.Ramana, [2] G Michael, [3] Dr.AR.Arunachalam
[1] PG Student Assistant Professor Department of CSE Bharath University Chennai-73.
[2] Assistant Professor Department of CSE Bharath University Chennai-73.
[3] Professor & Head Professor & Head Department of CSE Bharath University Chennai-73.

Abstract: ***Conventional spatial queries, such as range search and nearest neighbor retrieval, involve only conditions on objects based on different selection criteria.Today, 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. The important of spatial database is reflected by the convenience of modeling entities of reality in a geometric manner.Location of restaurants, hotels, hospitals and so on is often represented as points in a map.Currently, the best solution to such queries is based on the IR2-tree.we are interested in a more general form of local search, that is, to search local content on the Web. In our approach, each web page will be first assigned to a few geographical locations according to its content and then spatially indexed in the search engine.A straightforward approach is to treat geographical words which represent location information as common keywords, and to retrieve web pages with specified location names in the same way to keyword matching.To solve the problem, it is necessary to design an efficient index structure that considers both spatial and textual features of web pages. How to efficiently index and search location-specific information is being a key problem for location based search engines***

I. INTRODUCTION

Location-specific information is common on the Web. According to previous studies, nearly one fifth of web search tasks are related to a specific place or region which is usually called location based web search. Recently, more and more commercial search engines start to provide location based services, such as local search, local advertisements and map services. These services are particularly useful for mobile users. Most commercial search engines, such as Google Local and Yahoo! Local only search business addresses in Yellow Pages or other kinds of paid lists. In this paper, we are interested in a more general form of local search, that is, to search local content on the Web. In our approach, each web page will be first assigned to a few geographical locations according to its content and then spatially indexed in the search engine. Therefore, it can be later retrieved by its locations. How to efficiently index and search location-specific information is being a key problem for location based search engines. A straightforward approach is to treat geographical words which represent location information as common keywords, and to retrieve web pages with specified location names in the same way to keyword matching. However, simple keyword matching neglects underlying spatial relationships, therefore, does not support advanced spatial queries. To solve the problem, it is necessary to design an efficient index structure that considers both spatial and textual features of web pages.

II. LITERATURE SURVEY

The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles

The R-tree, one of the most popular access methods for rectangles, IS based on the heuristic optimization of the area of the enclosing rectangle m each inner node By running numerous experiments m a standardized tested under highly varying data, queries and operations, we were able to design the R*-tree which incorporates a combined optimization of area, margin and overlap of each enclosing rectangle m the directory Using our standardized tested m an exhaustive performance comparison, It turned out that the R*-tree clearly outperforms the existing R-tree varmntsGuttman's linear and quadratic R-tree and Greene's variant of the R-tree.This superiority of the R*-tree holds for different types of queries and operations, such as map overlay. for both rectangles and multidimensional points m all experiments From a practical point of view the R*-tree 1s very attractive because of the following two reasons 1 It efficiently supports point and spattal data at the same time and 2 Its implementation cost 1s only slightly higher than that of other R-trees.

In this paper we will consider spatial access methods (SAMs) which are based on the approximation of a complex spatial object by the minimum bounding rectangle with the sides of the rectangle parallel to the axes of the data space. The most important property of this simple approximation 1s that a complex object 1s represented by a limited number of bytes Although a lot of information 1s lost, minimum bounding rectangles of spatial objects preserve the most essential geometric properties of the object, 1 e the location of the object and the extension of the object in each axis In [SK 881 we showed that known SAMs organizing (minimum bounding) rectangles are based on an underlying point access method (PAM) using one of the following three techniques clipping, transformation and overlapping regions.

Retrieving Topk Prestige based Relevant Spatial Web Objects

The 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. The paper 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. Studies suggest that at least some 20% of all web queries have local intent, meaning that the queries target local content.

In step with the web being used increasingly by mobile users, this percentage can be expected to increase. Next, geo-positioning is increasingly available for mobile devices, e.g., by means of built-in GPS receivers. This enables web users who query for local content to provide their locations to services. Search engines already recognize local intent, and specialized services, e.g., maps and yellowpage services that target local content continue to proliferate. For example, travel sites such as TripAdvisor and TravellersPoint offer services that enable users to find hotels with particular facilities and located in particular regions. Several proposals already exist for the querying for geo-located web content, termed spatial web objects. A location-aware keyword querytakes a location and specified keywords as arguments and returns web objects that are ranked according to both spatial proximity and text relevance relative to the query. Some proposals view keywords as Boolean predicates, filtering out web objects that do not contain the keywords and ranking the remaining objects based on their spatial proximity to the query. Other proposals combine spatial proximity and textual relevance using a linear ranking function.

Efficient Retrieval of the Top Most Relevant Spatial Web Objects

The conventional Internet is acquiring a geo-spatial dimension. Web documents are being geo-tagged, and geo-referenced objects such as points of interest are being associated with descriptive text documents. The resulting fusion of geo-location and documents enables a new kind of top-k query that takes into account both location proximity and text relevancy.

To our knowledge, only naive techniques exist that are capable of computing a general web information retrieval query while also taking location into account. This paper proposes a new indexing framework for location aware top-k text retrieval. The framework leverages the inverted file for text retrieval and the R-tree for spatial proximity querying. Several indexing approaches are explored within the framework. The framework encompasses algorithms that utilize the proposedindexes for computing the top-k query, thus taking into account both text relevancy and location proximity to prune the search space. Results of empirical studies with an implementation of the framework demonstrate that the paper's proposal offers scalability and is capable of excellent performance. Driven in part by the emergence of the mobile Internet, the conventional Internet is acquiring a geo-spatial dimension. On the one hand, many (geo-referenced) points of interest—e.g., stores, tourist attractions, hotels, entertainment services, public transport, and public services—are being associated with descriptive text documents. On the other hand, web documents are increasingly being geo-tagged. This fusion of geo-location and documents enables queries that take into account both location proximity and text relevancy. One study has found that about one fifth of web search queries are geographical and have local intent, as determined by the presence of geographical terms such as place names and postal codes. Indeed commercial search engines have started to provide location based services, such as map services, local search, and local advertisements. For example, Google Maps supports location-aware text retrieval queries. Additional examples of location-based services include online yellow pages.

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. Existing works mainly focus on finding top-k Nearest Neighbours, where each node has to match the whole querying keywords .It does not consider the density of data objects in the spatial space. Also these methods are low efficient for incremental query.

PROPOSED SYSTEM

A spatial database manages multidimensional objects (such as points, rectangles, etc.), and provides fast access to those objects based on different selection criteria. The importance of spatial databases is reflected by the convenience of modelling entities of reality in a geometric manner. For example, locations of restaurants, hotels, hospitals and so on are often represented as points in a map, while larger extents such as parks, lakes, and landscapes often as a combination of rectangles. Many functionalities of a spatial database are useful in various ways in specific contexts. For instance, in a geography information system, range search can be deployed to find all restaurants in a certain area, while nearest neighbour retrieval can discover the restaurant closest to a given address.

MODULES DESCRIPTION

Data Station Broadcast Launch
Mobile Host Creation
Location Details Update
Man-In-Middle Generation
Forward Lbsq To Nearest Neighbour
Retrieving Lbsq And Data Filtering

DATA STATION BROADCAST LAUNCH

The Data Station Broadcast Launch module Used to start the Server. Its enable the Server Ip Address and port address to listening the client nodes. It also lists the active nodes currently in the Client side. Here the Admin only has the privileges to start the server. It will help to store the data in web server. This is important state of mobile environment.

MOBILE HOST CREATION

In this module the Admin going to enter the mobile host details such as Mobile Host Id, Mobile Host Name, Password and Location .It will used to connect the clients. In later these details are used to connect the appropriate client's database .this database information are used to access the data from web server and to maintain the details of accessing information.

LOCATION DETAILS UPDATE

In this module Admin update the Category, Category name, Address and location. These detail service are fully depends on the clients. Through this details other clients are going to search the location based queries. This is the source of location based queries to accessing the details from web server. These are all data accessed by more than one client from web server database.

MAN-IN-MIDDLE GENERATION

The Man-In-Middle creates estimated valid regions (EVRs) for mobile clients by exploiting spatial and temporal locality of spatial queries on NN query history and available data objects, respectively. The Man-In-Middle maintains an object

IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 3, Issue 4, April 2018

cache and two index structures: an EVR-tree for NN queries and a grid index for window queries. The two index structures share the data objects in the object cache. The EVR-tree is an R-tree (or its variants) composed of EVRs where each EVR is wrapped in a minimum bounding box (MBR). An EVR consists of the region vertices with respect to a data object and a pointer to the corresponding object entry in the object cache.

FORWARD LBSQ TO NEAREST NEIGHBOUR

The sending mobile host LBSQ module used to send the Location based spatial query to the Server. The Clients are the privileged person to send the Query to the server. They get the Response based on the Locations .These details are resided in the spatial database of the server. Here the web server receives more than one query from more than users to make overload. That time the Man-In-Middle server database will help to reduce the overload.

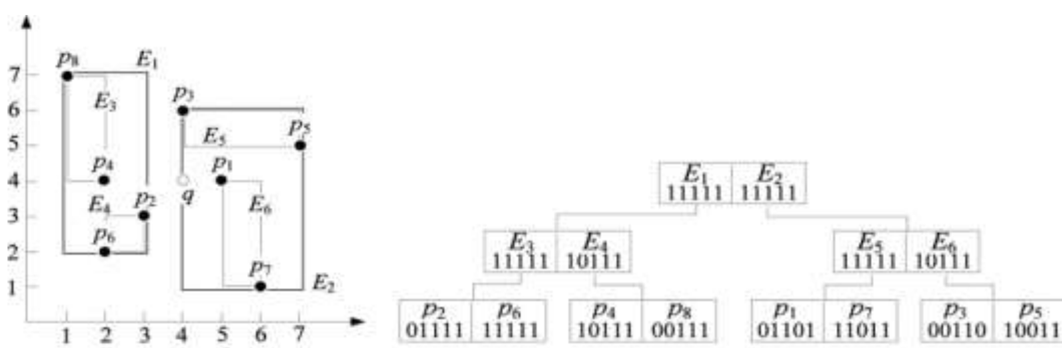
RETRIEVING LBSQ AND DATA FILTERING

In this module the client node search the nearest nodes to get the response if that node contains the particular query it will response to the corresponding client otherwise the query forwarded to the server. Then the server Filter details based on the query and those details are sending to the client. Finally the client side users retrieve the data without any distortion from the web server database with secured. Here the data accessing speed is increase.

ALGORITHM/METHOD SPECIFICATION

The IR2-tree, however, also inherits a drawback of sig-nature files: false hits. That is, a signature file, due to its conservative nature, may still direct the search to some objects, even though they do not have all the keywords. The penalty thus caused is the need to verify an object whose satisfying a query or not cannot be resolved using only its signature, but requires loading its full text description, which is expensive due to the resulting ran-dom accesses. It is noteworthy that the false hit problem is not specific only to signature files, but also exists in other methods for approximate set membership tests with compact storage (see [7] and the references therein). Therefore, the problem cannot be remedied by simply replacing signature file with any of those methods.

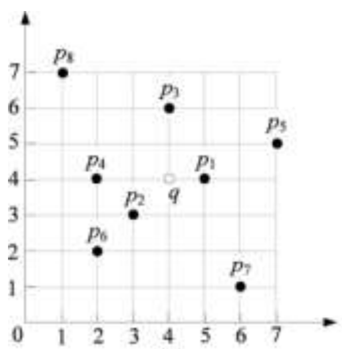
In this project, we design a variant of inverted index that is optimized for multidimensional points, and is thus named the spatial inverted index (SI-index). This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate 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. We can (sequentially) 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 demon-strated by experiments, the SI-index significantly outper-forms the IR2-tree in query efficiency, often by a factor of orders of magnitude.



(a) (b)

FUTURE ENHANCEMENT

we will continue to improve the performance for location indexing. Geographical ranking is also an important problem to study, which is critical for improving the performance of location based web search. In our approach, we represented the geographical scopes of web pages as multiple MBRs and compared three hybrid index structures based on inverted files and R*-trees. We have also developed a complete location based search engine and carried our large scale experiments to validate the proposed structures. Experiments showed the structure of first inverted file then R*-tree is the most efficient in query time.



(a)

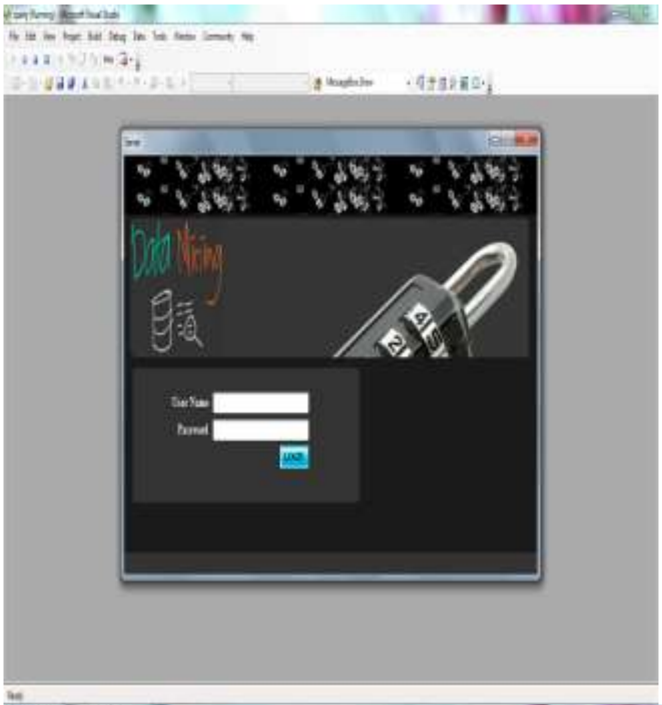
p	W_p
p_1	{a, b}
p_2	{b, d}
p_3	{d}
p_4	{a, e}
p_5	{c, e}
p_6	{c, d, e}
p_7	{b, e}
p_8	{c, d}

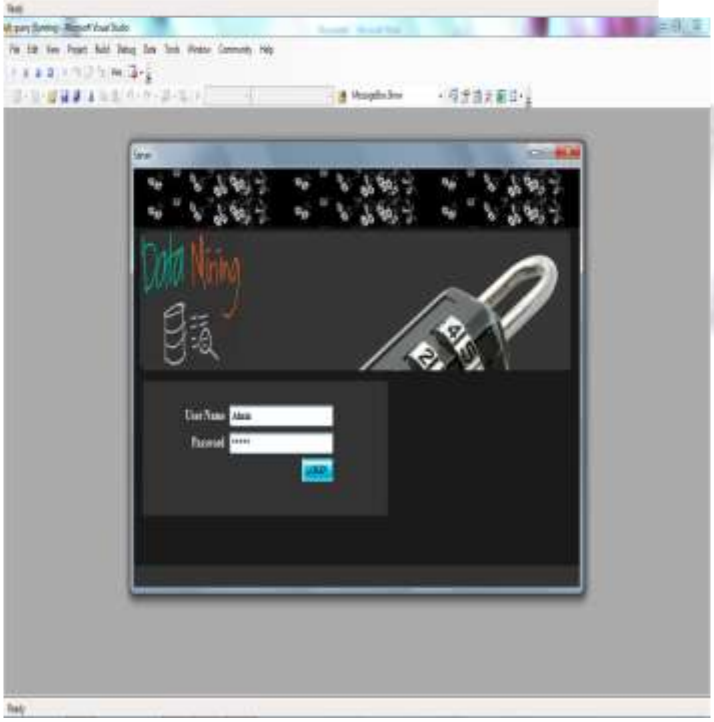
(b)

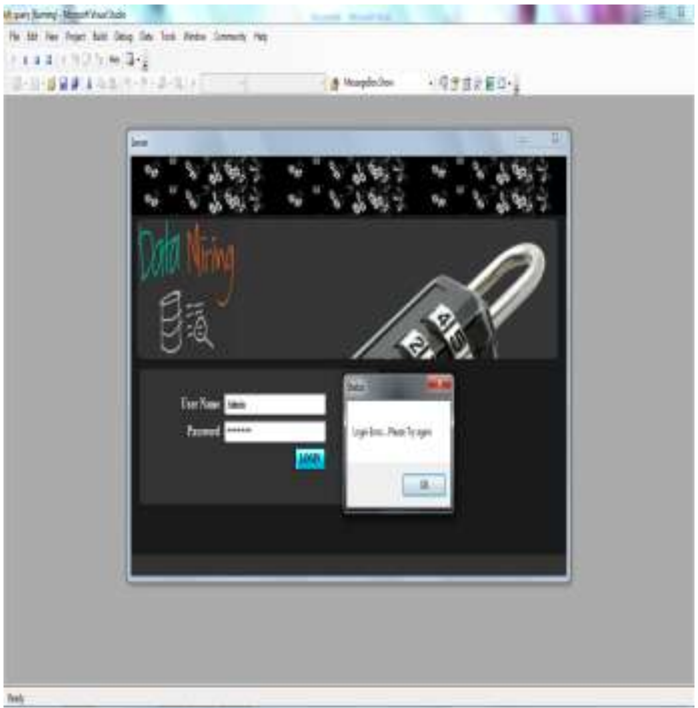
III. CONCLUSION

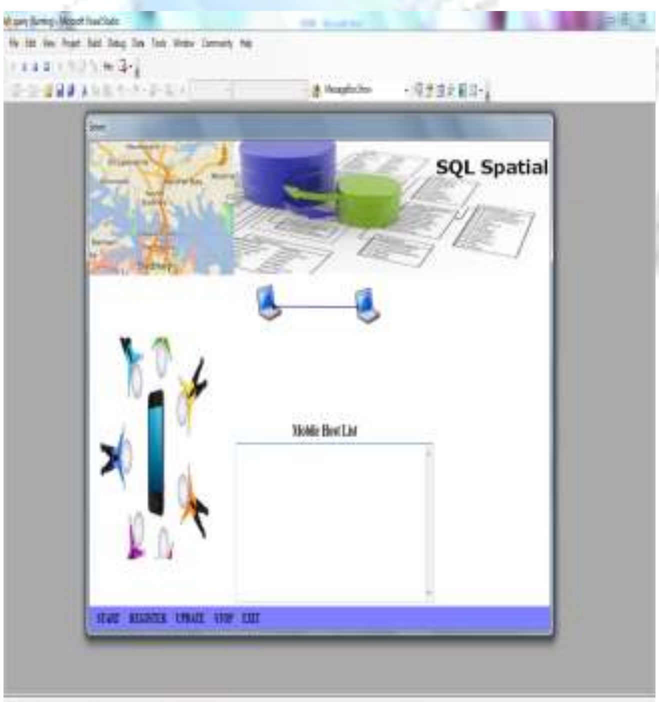
We have seen plenty of applications calling for a search engine that is able to efficiently support novel forms of spatial queries that are integrated with keyword search. The existing solutions to such queries either incur prohibitive space consumption or are unable to give real time answers. In this 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-inde 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.

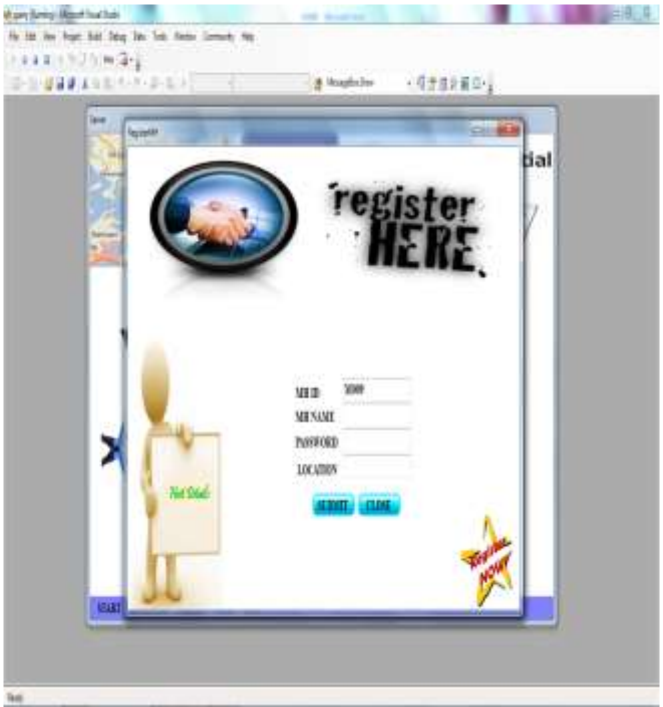
VI. SCREEN SHOTS

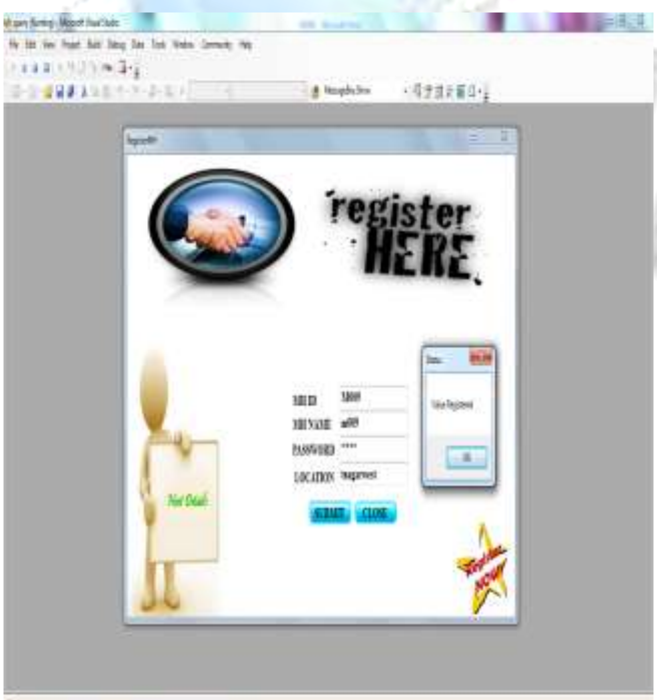




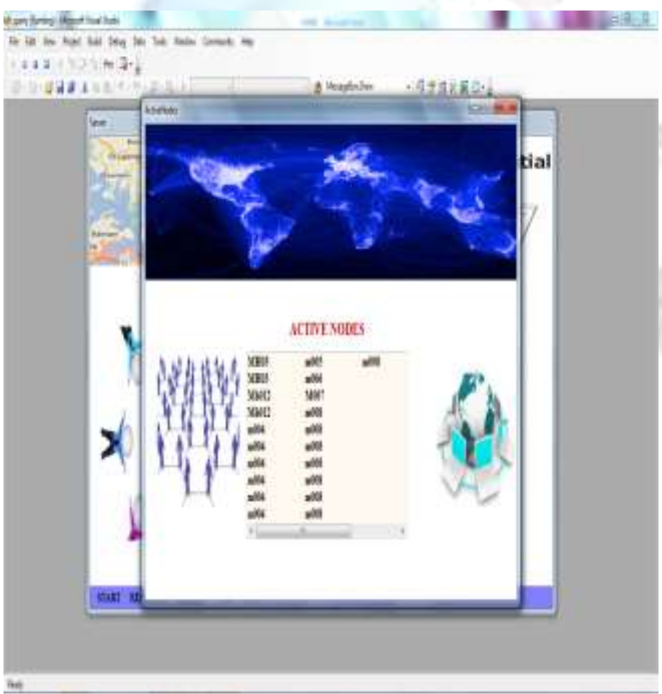

Data Mining
User Name
Password

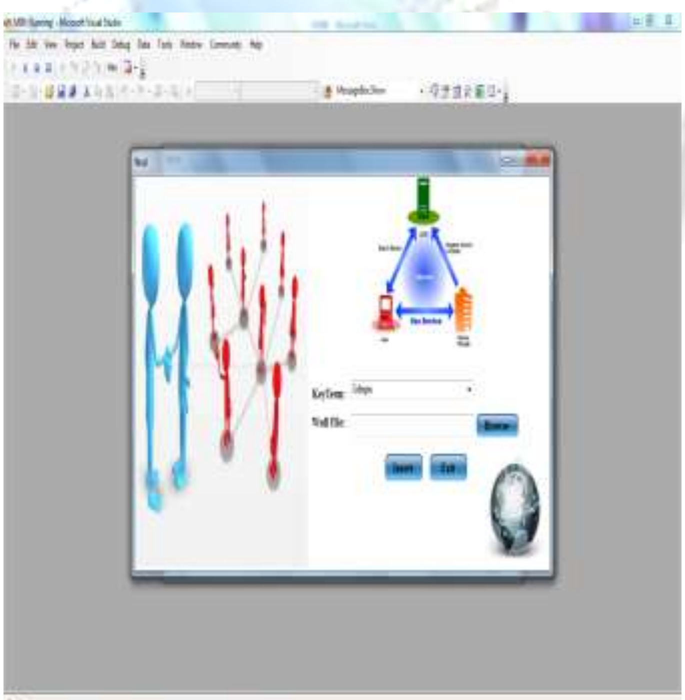

SQL Spatial

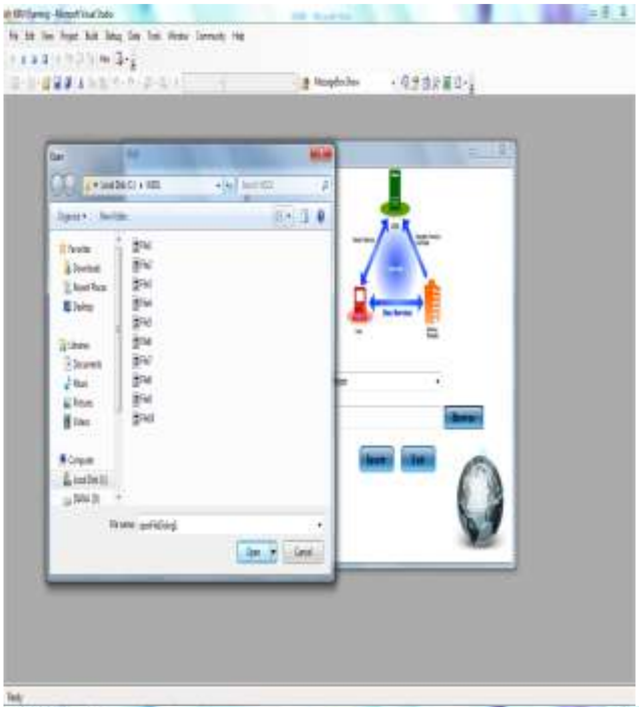

register
HERE

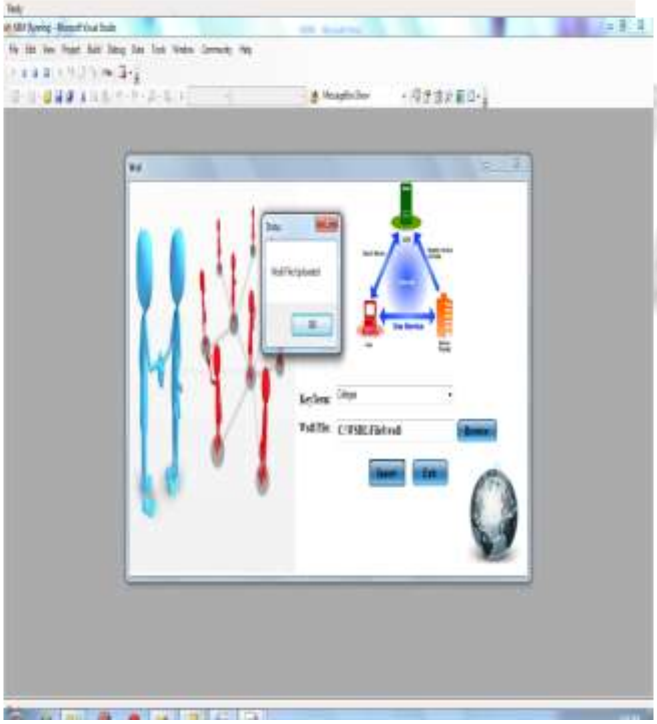

register
HERE

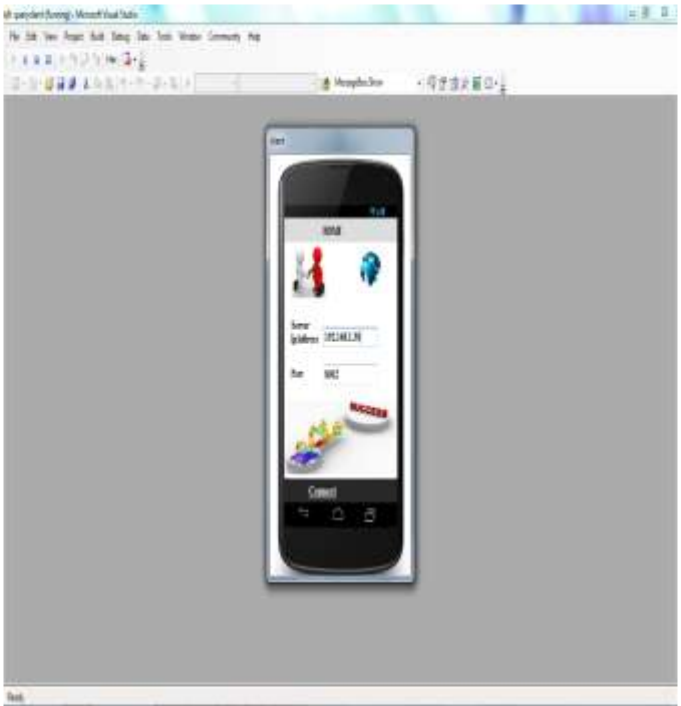

SNO
CATEGORY
NAME
LOCATION
ADDRESS
PHONE NO
ENTER
CLOSE


ACTIVE NODES








Connect


USER
LOGIN

REFERENCES

1. N.Beckmann, H. Kriegel, R. Schneider, and B. Seeger, "The R- tree: An Efficient and Robust Access Method for Points and Rectangles," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 322-331, 1990.
2. 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.
3. Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 277-288, 2006.
4. X.Cao, G. Cong, and C.S. Jensen, "Retrieving Top-k Prestige- Based
5. Relevant Spatial Web Objects," Proc. VLDB Endowment, vol. 3, no. 1, pp. 373-384, 2010.

6. V.Hristidis and Y. Papakonstantinou, "Discover: Keyword Search in Relational Databases," Proc. Very Large Data Bases (VLDB), pp. 670-681, 2002.
7. J.Lu, Y. Lu, and G. Cong, "Reverse Spatial and Textual k NearestNeighbor Search," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 349-360, 2011.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