Exploring Knowledge for a Common Man Through Mobile Services and Knowledge Discovery in Databases

Sanjeev Manchanda* smanchanda@thapar.edu
School of Mathematics and Computer Applications,
Thapar University, Patiala-147004 (INDIA)
**Corresponding author*

Mayank Dave mdave67@yahoo.com
Department of Computer Engg.,
National Instt. Of Technology, Kurukshetra, India

S. B. Singh sbsingh69@yahoo.com
Department of Mathematics,
Punjabi University, Patiala, India

ABSTRACT

Every common man needs some guidance/advice from his/her friends/relatives/known one, whenever he/she wants to buy something or want to visit somewhere. In this paper authors propose a system to fulfill the knowledge requirements of a common man through mobile services using data mining and knowledge discovery in databases. This system will enable us to furnish such information to a common man without any cost or at a low cost and with least effort. A comparison of proposed system is provided with reference to other available systems.

Keywords: Data mining, Knowledge Discovery in Databases, Global Positioning System, Geographical Information System, Mobile Services.

1. INTRODUCTION

Science and technology in its roots is meant for serving human community in masses. But latest trends are away from their original thought. If we look towards Indian scenario, we have started working for big organizations, which are meant for making big profits rather working for the up-lifting the society. This paper is an effort to help a common human being through automating knowledge discovery to contribute him/her in its day-to-day requirements. Information Technology and communication has facilitated a common person to connect one to whole world. It is one of the biggest achievements of science and technology. In continuation of same technology and incorporation of data mining and knowledge discovery in databases can help anyone to gain some knowledge through the information stored in databases scattered worldwide. This paper will furnish a base to fulfill the information/knowledge requirements of a common human being in its day-to-day proceedings. Present paper presents architecture of system to develop a sophisticated system to furnish customized information needs of users. Present paper also compares proposed system with other major competitive systems. This paper is organized as following section introduces the foundational concepts behind the development of proposed system. Third section contains literature review, Fourth section describes the problem, Fifth section briefly describes possible solution, Sixth section presents system's architecture including algorithm of the system, Seventh section discusses issues involved in developing and commercializing proposed system, Eighth section includes the description of technology used for development of system, Ninth section compares proposed system with other systems as well as other results, Tenth section discusses current state of the system, Eleventh section concludes the paper as well as discusses future aspects and last but not the least Twelfth section contains references used in present paper.

1 2. KDD, GPS, GIS and Spatial Databases

2

3 2.1 Data mining and knowledge discovery in databases

Knowledge discovery in databases (KDD) is a drift to search new patterns in existing databases through the use of data mining techniques. The techniques used earlier to search data in response to queries having fixed domains have changed to search unknown patterns or vaguely defined queries.

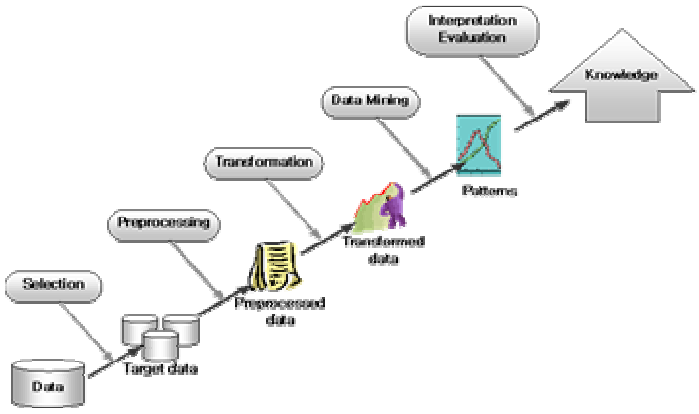


FIGURE 1: KDD Process

2.2 Geographic Information System

Geographic Information System (GIS) is a computer based information system used to digitally represent and analyse the geographic features present on the Earth' surface and the events

(non-spatial attributes linked to the geography under study) that taking place on it. The meaning to represent digitally is to convert analog (smooth line) into a digital form.

2.3 Global Position System

The Global Positioning System (GPS) is a burgeoning technology, which provides unequalled accuracy and flexibility of positioning for navigation, surveying and Geographical Information System data capture. The GPS NAVSTAR (Navigation Satellite timing and Ranging Global Positioning System) is a satellite-based navigation, timing and positioning system. The GPS provides continuous three-dimensional positioning 24 hrs a day throughout the world. The technology seems to be beneficiary to the GPS user community in terms of obtaining accurate data up to about100 meters for navigation, meter-level for mapping, and down to millimeter level for geodetic positioning. The GPS technology has tremendous amount of applications in GIS data collection, surveying, and mapping.

2.4 Spatial Database

A spatial database is defined as a collection of inter-related geospatial data, that can handle and maintain a large amount of data which is shareable between different GIS applications. Required functions of a spatial database are consistency with little or no redundancy, maintenance of data quality including updating, self descriptive with metadata, high performance by database, management system with database language and - security including access control

3. Literature Review

Recently, the data mining and spatial database system has been subjects of many articles in business and software magazines. However, two decade before, few people had heard of the terms data mining and spatial/geographical information system. Both of these terms are the evolution of fields with a long history, the terms themselves were only introduced relatively recently, in late 1980s. Data mining roots are traced back along three family lines viz. statistics, artificial intelligence and machine learning, whereas roots of Geographical Information System lies from the times when human being started traversing to explore new places. Data mining, in many ways, is fundamentally the adaptation of machine learning techniques to business applications. It is best described as the union of historical and recent developments in statistics, artificial intelligence and machine learning. These techniques are then used together to study data and find hidden trends or patterns within. Data mining and Geographical Information systems are finding increasing acceptance in science and business areas that need to analyze large amounts of data to discover trends, which they could not otherwise find. Combination of geography and data mining has generated a great need to explore new dimensions as Spatial Database Systems.

Formal foundation of data mining lies from a report on (International Joint Conference on Artificial Intelligence) IJCAI-89 Workshop (Piatetsky 1989), in which he emphasized on the need of data mining due to growth in amount of large databases. This report confirmed the recognition for the concept of Knowledge Discovery in Databases. Data mining term got its recognition from this report and became familiar in scientific community soon. Initial work was being done on classification and logic programming in 1991. Han et al., 1991 emphasized on concept based classification in Relational database. He devised a method for the classification of data in relational databases by concept-based generalization and proposed an algorithm concept-based data classification algorithm called DBCLASS, whereas Bocca, 1991 discussed about design and engineering of the enabling technologies for building Knowledge Based Management Systems (KBMS). He showed the problems with the required solutions of existing technologies commercially available i.e., relational database systems and logic programming. Aggarwal et al., 1992 discussed about the problems with classifying data populations and samples into m group intervals. They proposed a tree based interval classifier, which generates a classification function for each group that can be used to efficiently retrieve all instances of the specified group from the population database.

The term "spatial database system" has become popular during the last few years, to some extent through the *Symposium on Large Spatial Databases,* which has been held biannually

since 1989 (Buchmann et al. 1990, Giinther et al. 1991 and Abel et al. 1993). This term is associated with a view of a database as containing sets of objects in space rather than images or pictures of a space. Indeed, the requirements and techniques for dealing with objects in space that have identity and well-defined extents, locations, and relationships are rather different from those for dealing with raster images. It has therefore been suggested that two classes of systems, *spatial database systems* and *image database systems,* be clearly distinguished (Buchmann et al. 1989, Frank 1991). Image database systems may include analysis techniques to extract objects in space from images, and offer some spatial database functionality, but they are also prepared to store, manipulate, and retrieve raster images as discrete entities.

Han et al., 1994 tried to explore whether clustering methods had a role to play in spatial data mining. They developed a clustering algorithm called CLARANS, which is based on randomized search and two data mining algorithms Spatial Dominant Approach (SD) and Non-Spatial Dominant Approach (NSD) and showed the effectiveness of algorithms. Mannila et al., 1994 revised the solution to the problem raised by Agarwal Rakesh et al. 1993 and proposed an algorithm to improve the solution. Han et al., 1994 studied the construction of Multi Layered Databases (MLDBs) using generalization and knowledge discovery techniques and the application of MLDBs to cooperative/intelligent query answering in database systems. They proposed an MLDB model and examined in their study and showed the usefulness of MLDB in cooperative query answering, database browsing and query optimization. Holsheimer et al., 1994 surveyed the data mining researches of that time and presented the main underlying ideas behind data mining such as inductive learning, search strategies and knowledge representations used in data mine systems, also described important problems and suggested their solutions. Kawano et al., 1994 integrated knowledge sampling and active database techniques to discover interesting behaviors of dynamic environments and react intelligently to the environment changes. Their studies showed firstly that data sampling was necessary in the collection of information for regularity analysis and anomaly detection, Secondly knowledge discovery was important for generalizing low-level data to high-level information and detecting interesting patterns, Thirdly active database technology was essential for the real time reaction to the changes in real time environment and lastly the integration of three technologies forms a powerful tool for control and management of large dynamic environments in many applications.

Data Classification technique contributed by Han et al. 1995, Kohavi 1996, Micheline et al. 1997, Li et al. 2001 and others, Han et al., 1995, 1998 stated that their research covered a large wide spectrum of knowledge discovery, which included the study of knowledge discovery in relational, object oriented, deductive, spatial, active databases and global information systems and development of various kinds of knowledge discovery methods including attribute-oriented induction, progressive deepening for mining multiple level rules, meta-guided knowledge mining *etc.* and also studied algorithms for the techniques of data mining. Later he investigated the issues on generalization-based data mining in object-oriented databases in three aspects generalization of complex objects, class based generalization and extraction of different kinds of rules. Their studies showed that set of sophisticated generalization operators could be constructed for generalization of complex objects, dimension based class generalization mechanism could be developed for class based generalization and sophisticated rules formation method could be developed for extraction of different kinds of rules.

The development of specialized software for spatial data analysis has seen rapid growth since the lack of such tools was lamented in the late 1980s by Haining, 1989 and cited as a major impediment to the adoption and use of spatial statistics by GIS researchers. Initially, attention tended to focus on conceptual issues, such as how to integrate spatial statistical methods and a GIS environment (loosely vs. tightly coupled, embedded vs. modular, etc.), and which techniques would be most fruitfully included in such a framework. Familiar reviews of these issues are represented in, among others, Anselin et al., 2000, Goodchild et al. 1992, Fischer et al., (1993, 1996, 1997), Fotheringham et al., (1993, 1994). In geographical analysis by Monmonier, (1989) made operational Spider/Regard toolboxes of Haslett, Unwin and associates (Haslett et al. 1990, Unwin 1994). Several modern toolkits for exploratory spatial data analysis (ESDA), also incorporate dynamic linking and to a lesser extent brushing. Some of these rely on interaction with a GIS for the map component, such as the linked frameworks combining XGobi or XploRe with ArcView (Cook et al. 1996, Symanzik et al. 2000), the SAGE toolbox, which uses ArcInfo

(Wise et al., 2001), and the DynESDA extension for ArcView (Anselin, 2000), GeoDa's immediate predecessor. Linking in these implementations is constrained by the architecture of the GIS, which limits the linking process to a single map (in GeoDa, there is no limit on the number of linked maps). In this respect, GeoDa is similar to other freestanding modern implementations of ESDA, such as the cartographic data visualizer, or cdv (Dykes, 1997), GeoVISTA Studio (Takatsuka et al., 2002) and STARS (Rey et al., 2004). These all include functionality for dynamic linking, and to a lesser extent, brushing. They are built in open source programming environments, such as Tkl/Tk (cdv), Java (GeoVISTA Studio) or Python (STARS) and thus easily extensible and customizable. In contrast, GeoDa is (still) a closed box, but of these packages it provides the most extensive and flexible form of dynamic linking and brushing for both graphs and maps.

Common spatial autocorrelation statistics, such as Moran's I and even the Local Moran are increasingly part of spatial analysis software, ranging from CrimeStat (Levine, 2004), to the spdep and DCluster packages available on the open source Comprehensive R Archive Network (CRAN),3 as well as commercial packages, such as the spatial statistics toolbox of the forthcoming release of ArcGIS 9.0 (ESRI, 2004). Continuous space in spatial data was discussed and presented join less approach for mining spatial patterns (Yoo et al. 2006).

One major aspect of any such systems is user satisfaction. User satisfaction depends on many aspects like usability, Accuracy of product, information quality etc. Usability is one of the most important factors in the phases of designing up to selling a product. (Nielsen, 1993). But the efficiency of a product is influenced by the acceptance of the user. Usability is one basic step to acceptance and finally to efficiency of a product. A new approach is the "User Centred Design", UCD. The prototyping is described with the ISO standard 13407: "Human centred design process for interactive systems". The main mantras used here are "Know your user!" and "You aren't the user!" Both slogans describe the importance of the user.(Fröhlich et al., 2002) Concluding from the own experience as a user to other user groups is precarious and should be avoided. It is only possible to understand the user groups and the context of usage by careful analysis. (Hynek, 2002). The User Centred Design focuses on the user and their requirements from the beginning of achieving a product. "Usability goals should drive design. They can streamline the design process and shorten the design cycle." (Mayhew, 2002) Factors like reliability, compatibility, cost, and so on affect the user directly. Usability factors influence the decision of the user indirectly and can lead to subconscious decisions that are hardly traceable. The Information Quality (IQ) is the connector between Data Quality and the user. General definitions for IQ are "fitness for use" (Tayi et al., 1998), "meets information consumers needs" (Redman, 1996) or "user satisfaction" (Delone et al., 1992). This implies data that are relevant to their intended use and are of sufficient detail and quantity, with a high degree of accuracy and completeness, consistent with other sources, and presented in appropriate ways. Many criteria depend on each other and in this case not all criteria will be used. The information quality is a proposal to describe the relation between application, data, and the users. (Wang et al., 1999).

After a lot of development in this area, phenomenal success has been registered by the entry of world's best IT organizations like Google, Oracle, Yahoo, Microsoft etc. A lot of online services have been made available by these organizations like Google Maps, Google Earth, Google Mobile Maps, Yahoo Maps, Windows Live Local (MSN Virtual Earth) and mapquest.com (2007) etc., present study will compare its features with many of these services and will elaborate the comparative advantage of proposed system.

4. Problem Definition

Every man/woman in this world need some guidance or advice from its friends/relatives/known people about some purchase/visit/traveling, related expenditure, prices and path to be followed to reach the destination, best outlets/places *etc*. One tries its best to explore such information from its means. Reliability is always a matter of consideration in this regard; still one manages either to visit one's house or picks up telephone/mobile phone to enquire the required information. Now question arises, whether these services can't be automated? Answer is why not and up to certain extent such services have been automated, which may furnish information regarding targeted products/services/destinations e.g. on-line help-lines, Internet *etc*. On-line help is available to

facilitate anyone information regarding intended products/services. Search engines may help anyone to retrieve information from hypertext pages scattered all around the world. One can collect information in traces from such sources and can join them to get some knowledge from them. Again reliability is on stake. e. g. health help lines in any city are provided to furnish information regarding health services available in the city. One can enquire regarding these services available in the city, but many things aspects which one will not be in a position to clarify *i.e.* nearest service center, quality of service, comparable pricing, way to reach the destination, contact numbers *etc.* So problem can be stated that every person in this world seeks some guidance or advice for day-to-day purchases/servicing/traveling and some service, which can furnish such service on demand, can be an interesting solution.

5.Proposed Solution

An automated solution to such day-to-day problem can be formed as follows:

Consider a scenario, when a common man picks up its mobile device (cell phone or laptop *etc.*) and types a short message and sends it to a knowledge service center number and in return gets a bundle of information within moments through responding messages.

Now we'll see how this can be made possible through the merger of mobile services and data mining. In previous sections we have explained the concept of data mining and knowledge discovery in databases. Now we'll present a system to find the solution of above discussed problem. In last we'll look for into issues involved in implementation of this solution. In last the conclusion of the paper will be presented.

4

5 6. System Architecture

We present a system that can help in finding solution to above discussed problem.

Higher level view of this system is just like any other Client/Server architecture connected through Internet or a Mobile Network.

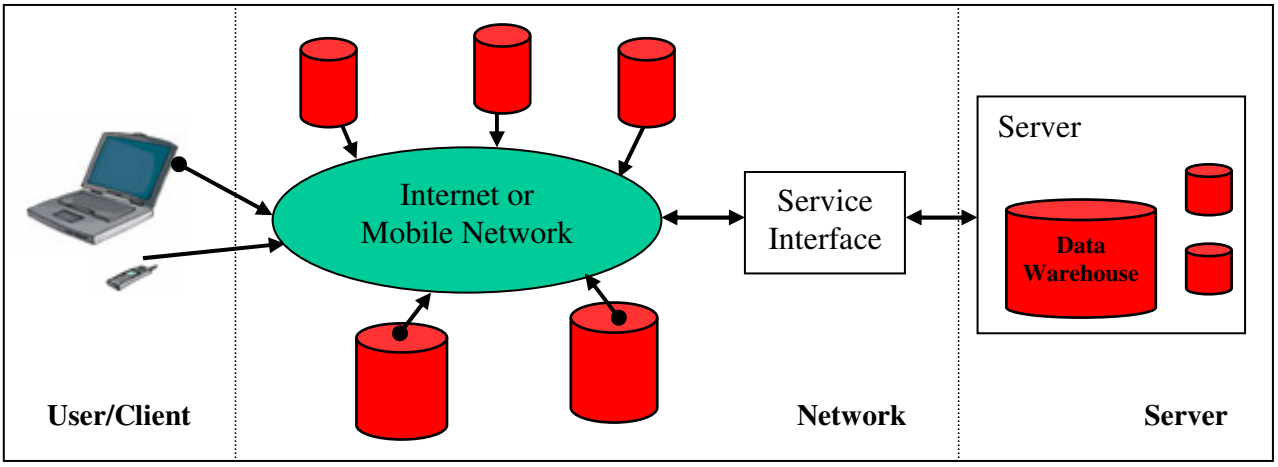


FIGURE 2: Client Server Architecture of the system

Client sends a message to Server through Internet or its mobile network. Network transfers this message further to server through a service interface.Service interface is connected to server through three interfaces. Depending upon the content of the message and forwards it any of the three interfaces of the server. If content has been received from a mobile phone or laptop/computer and its real time position/location is known, then message is forwarded to Client/Target Location Resolution Module, else if a mobile has sent a message without any information about its location then message is forwarded to Mobile Location Tracking Module, so that mobile's real time location can be identified else if a laptop/computer has sent a message without any information about its location then message is forwarded to Geographic IP Locator

Module , so that mobile's real time location can be identified. If message is forwarded to either Mobile Location Tracking Module or Geographic IP Locator Module, then after finding sender's current location message is forwarded to Client/Target Location Resolution Module to find the client's current and targeted location from spatial data warehouse. Incoming message may be in the form of an SMS (Short Message Service) from mobile or in the form of a query if obtained from a computer/laptop, so there may be need to convert the message into a query, where Query Conversion and Optimization module will help the system to fill the gap between actual message and internal representation of query. After converting message into suitable query Client/Target Location Resolution Module takes help from other modules to resolve the query. Processing of this query will use following algorithm.

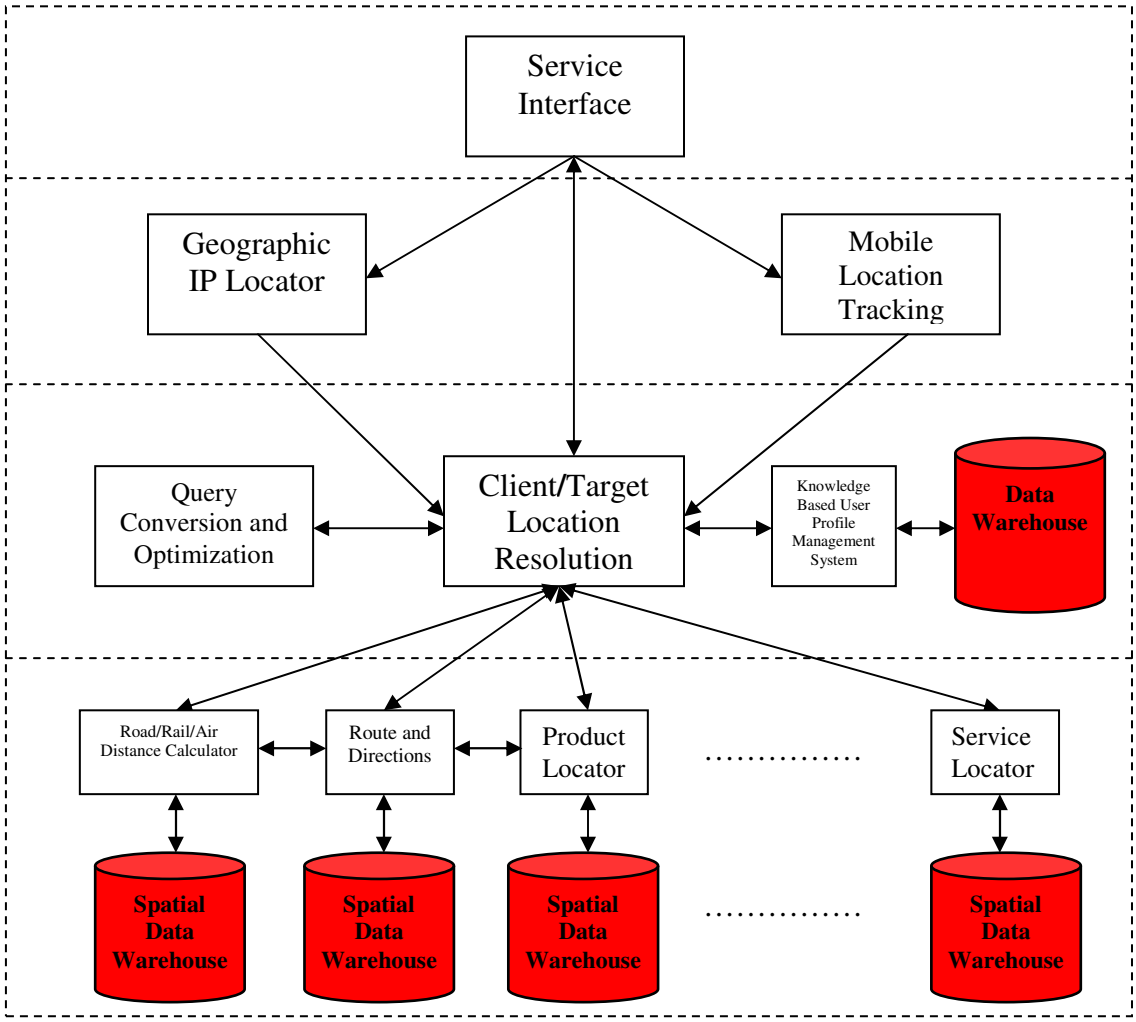


FIGURE 3: Server's Architecture

We shall discuss this algorithm in three aspects i. e. input, processing and output

6.1 Input

Input for this algorithm will be a message typed as a query in somewhat following formats or some similar representation that may need to be converted into SQL query.

product television, mcl india.punjab.patiala.tu, size '300 ltrs'; (1)

or

service travel, mtl india.punjab.patiala.bus_terminal; (2)

or

service food, quality best, mcl india.punjab.patiala.bus_terminal; (3)

[mcl/mtl stands for my (user's) current/target location]

This format includes target product or service, user's current location as well as extra information, which may be furnished on user's discretion, this extra information may involve many parameters, which may require to be standardized or may be defined as per algorithms implementation. Processing of Input from user will decide the response, so right kind of query increases more suitable response. After typing such message user sends it message to a knowledge service center, which initiates a process in response. Target of the process will be to return at least one success and at most a threshold number of successful responses.

6.2 Processing

Different modules will be initiated in response to the input. We discuss major aspects of resolving the problems related to search required information in databases.

6.2.1 User's Target Location Resolution

With the help of Mobile Location Tracking Module or Geographic IP Locator Module, we can identify client's current location. Many a times client's target location will be included in message itself, if we see at query number (2), then we find that client's targeted location is specified and finding such a location in spatial database is quite easy if this location is already included in database. Problem becomes complicated if client's target location is to be determined by the system itself e. g. in queries (1) and (3), for which a search process is initiated to find nearest neighbor that fulfills the demanded product path. Following modules within Client/Target Location Resolution Module will be initiated.

Module 1 First of all responding system searches for the user's current location and targets the place to search the solution. Whole spatial database is divided into four levels i. e. place level (Lowest Level i.e. to find target location at adjacent places), city level, state level and country level (Highest Level i. e. to find target location at adjacent countries). The user will define the level at which solution is required, if user enquire at some particular place level or city level or state level or national level, the search will also be implemented at same level, e.g. current search is implemented at place level within a city, search space will be adjacency of different places, if enquiry is at city level, search space will be at city level, algorithm will search in adjacent cities and so on up to national level.

Location will be targeted hierarchically as if a Domain Name System works as follows:

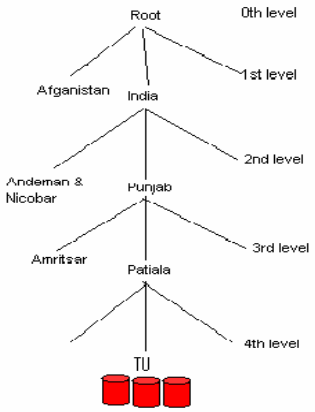


FIGURE 4: User's Current Location Identification.

User's requested search domain will search the databases at fourth level and will try to find the solution within the place first of all. If search finds related records from the databases of same place e.g. TU related records have information regarding the query, then system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the in user's desired format, otherwise search will involve next module.

Module 2 In this module a system search for the enquiry related records of the adjacent places and continues the search until all the adjacent places aren't searched for the solution of the query. Then system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and this module will continue search until all the adjacent places are not searched. If system is unable to retrieve the required knowledge, then it will involve the next module otherwise output will be furnished to the user in desired format.

FIGURE 5: Search within User's Current City i.e adjacent places.

Module 3 As this module is being involved, when the search has failed within the city, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 3rd level and starts finding solution of enquiry at state level by retrieving the records of adjacent cities and searches for the related records in the databases of these cities. This search continues until the success is not achieved by finding the records of related queries in adjacent cities and continues its expansion until databases of all the cities of the state aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will involve the next module.

FIGURE 6: Search for databases in adjacent cities within user's current state.

Module 4 As this module is being involved, when the search has failed within the state, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 2nd level and starts finding solution of enquiry at national level by retrieving the records of adjacent states and searches for the related records in the databases of these states. This search continues until the success is not achieved by finding the records of related queries in adjacent states and continues its expansion until databases of all the states of the country aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will involve the next module.

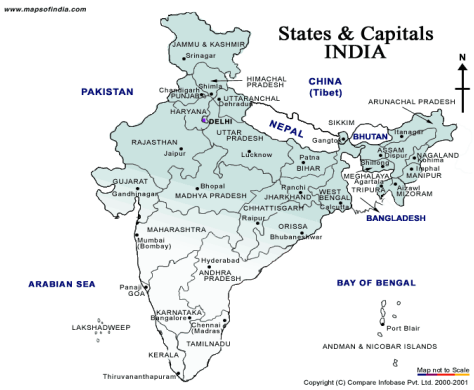


Figure 7: Search for databases at National Level.

Module 5 As this module is being involved, when the search has failed within the country, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 1st level and starts finding solution of enquiry at international level by retrieving the records of adjacent countries and searches for the related records in the databases of these countries. This search continues until the success is not achieved by finding the records of related queries in adjacent countries and continues its expansion until databases of all the countries of the world aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will report that the related information isn't available worldwide.

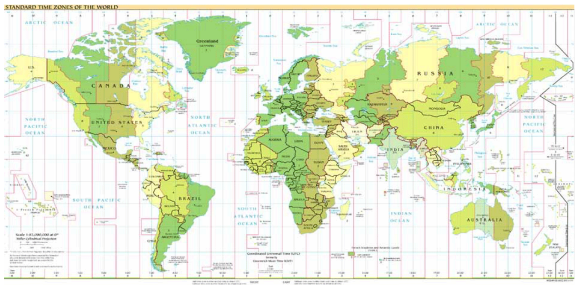


FIGURE 8: Search for databases at International Level.

Above discussed modules are to be implemented recursively, if we view the last three modules, one will observe that all three are doing the same thing then what is the necessity of having three different modules, answer lies in the sophistication involved in different levels, so these modules are required to be maintained separately or recursively in such a way that these can handle the complexity involved.

After fetching the records related to queries, task is to convert this information into useful knowledge for the user. Till now information collected is regarding product/service and the path to be traversed to reach the destinations. Now system will add its intelligence through its learned knowledge about the usefulness of the information, it will depend upon the kind of user of service is to be served. In this regard categories will be registered users and occasional users. System will serve both kind of users, but the implementation will differ as registered users will be furnished information on the basis of its profile and habit through continuously learning from the behaviour and choices of the user, whereas occasional users will be served that information, which the system will optimize to be important for the user. In this way it will optimize the information to be furnished for the user. Then system will calculate the distances total costs etc. to furnish the user important information.

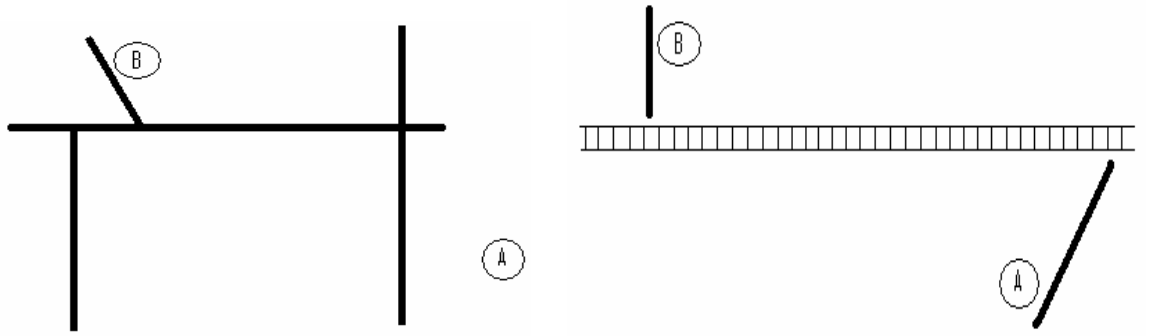
6.2.2 Calculation of Distance between Client's Current and Targeted location

After identifying Client's Current and Targeted locations it is obvious to calculate the distances between these locations. Usually any Global Positioning System is used to calculate Air distances of locations, it is due to the reason that Air distances can easily be calculated by finding the distance between pixels of these locations and multiplying it by scale used for differentiating two neighboring pixels, but calculating road and rail distances is very complex task. For a system to be practical for human beings, Air distances are not sufficient, Rail and Road distances need to be calculated to ease the users of the system. We have divided calculating distances into three parts i. e. Air, Rail and Road. Air distance calculating modules calculates Air as well as Water distances, whereas Rail and Road distance calculating module calculate their respective distances individually or as a combination of both. Calculating individual distances for Rail and Road distances is obvious, whereas combination of both is required when user is presently at a place away from railway station and it has to travel certain path through road. So client's total distance from its current location to targeted location will be a combination road and rail travel. Euclidean spaces are the most important metric for calculating distances between locations. We can define Euclidean space for arbitrary dimension n, but usually n=2, 3 is common in spatial information system as world can be defined in two or three dimensions. Thus, n-dimensional Euclidean space is defined as (R^n, Δ), where

$$R^n = \{(x_1, x_2, ..., x_n) : x_i \in R,\ \forall 1 \le i \le n\} \text{ and } \Delta(X,Y) = \left\{ \sum_{i=1}^{n} (x_i - y_i)^2 \right\}^{\frac{1}{2}},\ \forall X, Y \in R^n.$$

Most of the systems use Euclidean distance for calculating distances between two coordinates that is their Air distance. Distance between two pixels can be calculated quite easily, but situation becomes more complex when distances are not straight forward e. g. for Rail and Road.

Database can be maintained to map and to calculate the distances on Roads and Rail tracks or are made available by the authorities of respective countries, so that distances between two points on a road can be calculated. But problem appears with client's location or target location where no road or rail track is available. For example in Figure 9 a person wants to move from location 'A' to location 'B'. Air distances in both situation is lesser than the required path to be travelled by that person. In situation (a) person has to travel a distance that is a sum of distance from point 'A' to nearest road and through that road up to point 'B', whereas in situation (b) person has to travel distance from 'A' to rail track, through rail track up to a point near to 'B' on rail track and from that point through road up to point 'B'. Distance calculating module of proposed system calculates all these distances through Euclidean distance formula for different sub intervals of total that is from point 'A' to nearest road, then through road up to 'B' etc.



(a) Road travel distance problem **(b) Road and Rail travel distance problem**

FIGURE 9: Two situations where a person wants to move from A to B.

6.2.3 Databases required for acquiring information

After finding Client's Current/Target locations and calculating distances between them, question arises about the kind of products and services that can be provided to a user and responses provided to users about their queries need to be customized according to their preferences. For example if a customer is looking for a hotel to have meal, then responses to such queries must be based upon knowledge extracted from the past experiences of system with users, distance cannot be the only criteria for finding a solution and system must be equipped with some knowledge based implementation that can handle such aspects. We have equipped our system with a knowledge based implementation for registered customers. Registered customers provide their profile and behavior of registered people is monitored by the system, so that suitable responses can be provided facilitated to them. System includes different spatial databases for different kind of products and services as well as databases containing registered users' profiles.

If we view whole process broadly, then two types of major subsystems are combined to furnish the important information to the user; first one is enhanced Global Positioning System, which will help in targeting users' locations as well as location targeted by users and calculating distances between these locations, whereas second one is Knowledge Discovery System which will help in managing profiles of users, retrieving and facilitating information to users by adding its intelligence to process the information.

6.3 Output of the Queries

Output of the queries discussed in section 6.1 will be somewhat in following format, it contains information in map as well as textual format and this format has potential to be changed as per the enhancements in state-of-the-art technologies. Directions are provided visually as well as in textual format.

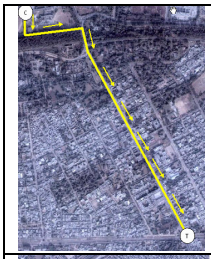
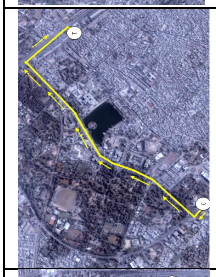
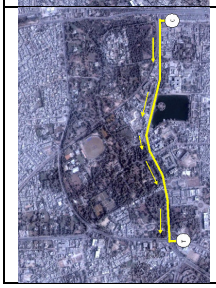
	Product Television Available at 22 No Railway crossing Route TU → 22 No Railway Crossing Price Wide Range 10K-25K Distance 1.5 Kms Contact No. 91-175-2390000,0175-2390001 Links www.abc.com
	Travel 'Bus Terminal' Current Location 'Fountain Square' Route Fountain Square → SBP Square → Railway Station → Bus Terminal Distance 3.25 Kms Enquiry Number 91-175-2390000
	Service Food Quality Best Target Location XYZ Restaurant, Fountain Square Route Bus Terminal → SBP Square → Fountain Square Distance 3 Kms Contact No. 91-175-2390002,0175-2390004 Links www.xyz.com

FIGURE 10: Output of queries on mobile devices.

In this way information can be furnished to the user on its mobile device with proper directions and information.

7. Issues in Implementing Knowledge Services to user through Mobile Devices

There are certain issues involved in commercializing proposed system.

7.1 Most important issue is regarding availability of such databases, their access, their connectivity, continuous updating and their schema. There is requirement to make such databases available on-line in some standard way, so that information can be exchanged and produced on demand. Governments maintain such databases; telephone exchanges, city help-lines, yellow pages services *etc*. can be of great use. But standardization will be a big trouble in this regard, which can be facilitated by implementing XML standards.

7.2 Whole world is required to be divided into identifiable places as places form cities, cities form states, states form countries and countries form the world. If required one more level can be incorporated by dividing cities into sub-cities and sub-cities into places. It will be very cumbersome task to implement such a thing and name as well as overlapping areas' conflict is also involved.

7.3 Another issue is regarding initiation of search, when system starts searching in other city, state or country, from which place, city or state respectively it should start searching. One possible solution may be to start the search from centralized place e.g. bus stand or railway stations for cities and capitals for states and countries *etc*.

7.4 Development of efficient algorithms is also an issue of concern.

7.5 Another issue is regarding knowledge center regulating authorities, commercial authorities can't be trusted to furnish information impartially.

7.6 Security issue and prompt delivery of service is pre-requisite in any on-line service.

7.7 Another issue is regarding Input/Output format standardization. Lack of standardization will not make this concept popular and individually implemented protocols will create much bigger problems.

7.8 Another issue is regarding privacy of the users; it may be the case information made available to such a service can be misused easily, so privacy of user is also an issue of concern.

6 8. Implementation

Above discussed algorithm is implemented in prototype of the system using J2EE 5.0 and Java 2 Platform Micro Edition (J2ME) is implemented and a database (using different packages like Oracle 10g, SQL Server etc.) is made available for implementing the above discussed system. The results have been obtained from the queries being invited by the campus students and the results are tested in highly simulated environment by observing different user selecting the outcomes. Maximum number of responses to be provided by system to user for each query was fixed at five. On the basis of such queries different results were being collected and it was being reviewed pilot testing and different responders are asked to value different results. On the basis of different queries and testing the response on the basis of some objective, descriptive criteria and satisfaction from responses of different queries have been given the percentage value to response and following experiments have been conducted within the campus.

7

8 9. Comparative and Experimental Results

Around fifty queries have been fed into the system to test the learning of systems on repetition of similar queries. System responds to the queries is on the basis of least distance, but responses are analyzed on the basis of user's satisfaction from responses as different users may prefer different responses and their review on the output is being given weightage and their choices are fed into system as per their behavior on responses, which is being noted by the system automatically. This prototype of the system is repeated for the processing of queries and results are observed, which are described here. First we present a comparative study of proposed system and other systems available yet, then we analyze the learning aspect of system.

9.1 Comparison of Proposed System with other Systems

If only mobile devices based such systems are considered then Google Mobile Maps is available and considering internet based such systems we can find systems that there are many systems like Google Maps, Google Earth, Yahoo Maps, Windows Live Local (MSN Virtual Earth) and Mapquest.com etc. are widely being made available. First biggest difference between proposed system and these systems is that no other system provides a client oriented service, only general information is provided, whereas proposed system is designed to furnish customized services to clients. Second difference is with respect to calculation of distances, as only Air distances are calculated by other systems, whereas proposed system will calculate actual distance. Google has started calculating real distances, but combination of different transportation modes, their distances, charges etc. are defined by proposed system, which is unavailable in any system present worldwide. Third foremost difference is the kind of services to be provided to users, Google earth can be considered as capable enough to be compared, as it is providing information related to many products and services, but problem with these services is that only limited number of options are made available, whereas proposed system is designed to involve exhaustive list of products and services as well as large number of options is planned to be included so that every human being can be benefitted through this. Fourth major difference is facilitating directions to users to reach up to destination, MSN Virtual Earth is providing this

feature at a small level, but proposed system is designed to furnish directions extensively so that user may be guided appropriately. Proposed system will go beyond these capabilities and will provide a lot more information about target product/service like telephone/mobile number, website link, travelling/communication modes available etc. Complexity of proposed system is also very high as is for most of other systems. Above comparison is included in Table 1. This table is prepared through general observations and expert opinion about different systems. Table 1 provides an overview of comparative advantage of proposed system over other systems. Analysis of table 1 indicates that Google is close competitor of proposed system, still proposed is comparatively much ahead of others.

System / Dimensions of Comparison	Google Maps	Google Earth	Google Mobile Maps	Map Quest	Windows Live Local (MSN Virtual Earth)	Yahoo Maps	Proposed System
Real Distances	Moderate	High	Moderate	Low	Moderate	Low	High
Product Availability	Moderate	High	Low	Low	Low	Moderate	High
Customized Service	Moderate	Moderate	Moderate	Low	Low	Moderate	High
Combination of Multiple Transportation Modes	Moderate	Moderate	Low	Low	Low	Low	High
Directions	Low	Moderate	Moderate	Low	Moderate	Moderate	High
Targeted Users	Web	Web	Mobile	Web	Web	Web	Web/Mobile
Complexity of System	High	High	High	Moderate	Moderate	Moderate	High

Table 1: Comparison of Proposed system with other major competitive systems

With above comparison there are certain unique features with proposed systems as well as other system. For example Google has started providing real time traffic information, whereas in proposed system can find real time location of web user, which is unavailable to any other system yet.

9 9.2 Performance

System prototype is tested for upper threshold of responses to be three and five responses and the results are analyzed in the light of many parameters involving user's preferences and others. These parameters on the basis of different weights assigned to them are evaluated and following graphs are being prepared.

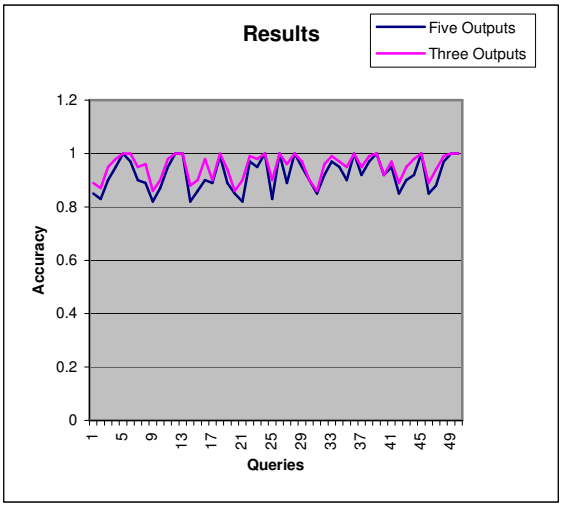


FIGURE 11: Accuracy of 50 Queries with Accuracy

FIGURE 12: Repetition of processing and Results

On analyzing the results of the queries and their accuracy in figures 11 and 12, we found that accuracy of results varies from 70% to 100% for five results for a single query and it varies from 82% to 100% for three outputs. As experiments were being conducted on a sample size of 50 queries one after the other and the number of responses were being kept at five and three. Decreasing the number of responses improved the results significantly. Also the learning process helped the system to improve the results significantly, when the same queries were being implemented on repetition system responded with an accuracy varying 82% to 100% for five outputs and an accuracy of 86% to 100% being observed. More experiments are undergoing to test the system with large number of queries and more repetition and more and more accuracy is expected to achieve in the range of 95% to 100%. With these results we are hopeful for the system to respond accurately in beta testing of prototype of system.

10

11 10. Work in Progress

One of the most important challenges before implementing this system is the challenges of identifying databases scattered worldwide for the search and their integration as different databases may differ in their configurations. So this prototype is undergoing the process of data integration by connecting different database standards available in the market. There is need to achieve an accuracy of 98% to 100%, so there is a need for strengthening the learning process. More learning processes are under development. Here decision tree is being implemented for searching the outputs. Other techniques (or hybrids) are also under consideration. For implementation of such a system public awareness is highly required, so that real implementation can catch up and the utility of system can be understood well by users. There is a need for marketing the system in this regard. This system will be implemented on several multiprocessing machines and server to server transactions will be experimented on them.

11. Future Expectations and Conclusion

We are very hopeful for the accomplishment of this system at Authors' affiliated place and future researches as well as standardization for its implementation and its acceptance worldwide. Such system may contribute for the up-liftment of society and also to bring science and technology to masses at large, So that a common person can be benefited through it.

12 12.Bibliography

1. Abel, D.J. and Ooi, B.C.(1993). An Introduction to Spatial Database Systems. Proceedings of the Third International Symposium on Large Spatial Databases, Singapore.
2. Aggarwal Rakesh, Ghosh Sakti, ImielinskiTomasz, Iyer Bala and Swami Arun (1992). An interval classifier for database mining applications. VLDB.
3. Agrawal Rakesh, Imielinski Tomasz and Swami Arun N. (1993). Mining Association Rules between Sets of Items in Large Databases. SIGMOD Conference: 207-216.
4. Anselin, L. (2000). Computing environments for spatial data analysis. Journal of Geographical Systems, 2(3):201–220.
5. Buchmann A. and Giinther O. (1990). Research issues in spatial databases. ACMSIGMOD Record, 19:61-68.
6. Buchmann A., Gfinther O., Smith T.R., and Wang Y.E. (1989), An Introduction to Spatial Database Systems. Proceedings of the First International Symposium on Large Spatial Databases, Santa Barbara,.
7. Cook, D., Majure, J., Symanzik, J., and Cressie, N. (1996). Dynamic graphics in a GIS: A platform for analyzing and exploring multivariate spatial data. Computational Statistics, 11:467–480.
8. Delone, W. H. and E. R. McLean (1992). "Information systems success: the quest for the dependent variable." Information Systems Research 3(1): 60-95.
9. Dykes, J. A. (1997). Exploring spatial data representation with dynamic graphics. Computers and Geosciences, 23:345–370.
10. ESRI (2004). An Overview of the Spatial Statistics Toolbox. ArcGIS 9.0 Online Help System (ArcGIS 9.0 Desktop, Release 9.0, June 2004). Environmental Systems Research Institute, Redlands, CA.
11. Fischer, M. and Nijkamp, P. (1993). Geographic Information Systems, Spatial Modelling and Policy Evaluation. Springer-Verlag, Berlin.
12. Fischer, M. M. and Getis, A. (1997). Recent Development in Spatial Analysis. Springer-Verlag, Berlin.
13. Fischer, M. M., Scholten, H. J., and Unwin, D. (1996). Spatial Analytical Perspectives on GIS. Taylor and Francis, London.
14. Fotheringham, A. S. and Rogerson, P. (1993). GIS and spatial analytical problems. International Journal of Geographical Information Systems, 7:3–19.
15. Fotheringham, A. S. and Rogerson, P. (1994). Spatial Analysis and GIS. Taylor and Francis, London.
16. Frank A. (1991). Properties of geographic data: Requirements for spatial access methods. Proceedings of the Second International Symposium on Large Spatial Databases, Zurich.
17. Fröhlich, M. and J. Mühlig (2002). Usability in der Konzeption. Usability Nutzerfreundiches Web-Design. Berlin.
18. Giinther, O. and Schek, H.-J. (1991). An Introduction to Spatial Database Systems. Proceedings of the Second International Symposium on Large Spatial Databases, Zurich.

19. Goodchild, M. F., Haining, R. P., Wise, S., and others (1992). Integrating GIS and spatial analysis — problems and possibilities. International Journal of Geographical Information Systems, 6:407–423.

20. Google Earth - Home(2007): http://earth.google.com

21. Google Maps(2007). http://maps.google.com

22. Google Mobile Maps(2007). http://www.google.com/gmm/index.html

23. Haining, R. (1989). Geography and spatial statistics: Current positions, future developments. In Macmillan, B., editor, Remodelling Geography, pages 191–203. Basil Blackwell, Oxford.

24. Han J., Fu Y., and Ng R. (1994). Cooperative Query Answering Using Multiple Layered Databases, Proc. 2nd Int'l Conf. on Cooperative Information Systems (CoopIS'94), Toronto, Canada, May, pp. 47-58.

25. Han J., Fu Y., Koperski K., Melli G., Wang W., Zaïane O. R. (1995). Knowledge Mining in Databases: An Integration of Machine Learning Methodologies with Database Technologies, Canadian Artificial Intelligence, October.

26. Han J., Nishio S., Kawano H., and Wang W. (1998). Generalization-Based Data Mining in Object-Oriented Databases Using an Object-Cube Model. Data and Knowledge Engineering , 25(1-2):55-97.

27. Han Jiawei, Cai Yandong and Cercone Nick (1991). Concept-Based Data Classification in Relational Databases. Workshop Notes of 1991 AAAI Workshop on Knowledge Discovery in Databases (KDD'91), Anaheim, CA, July, pp. 77-94.

28. Haslett, J., Wills, G., and Unwin, A. (1990). SPIDER — an interactive statistical tool for the analysis of spatially distributed data. International Journal of Geographic Information Systems, 4:285–296.

29. Holsheimer Marcel and Siebes Arno (1994). Data Mining: The Search for Knowledge in Databases. Technical Report CS-R9406, CWI, Amsterdam.

30. Hynek, T. (2002). User experience research- treibende Kraft der Designstrategie. Usability Nutzerfreundliches Webdesign. Berlin, Springer Verlag: 43-59.

31. Jorge B. Bocca (1991). Logic Programming Environments for Large Knowledge Bases: A Practical Perspective . VLDB: 563.

32. Kawano H., Nishio S., Han J., and Hasegawa T. (1994). How Does Knowledge Discovery Cooperate with Active Database Techniques in Controlling Dynamic Environment?, in Proc. 5th Int'l Conf. on Database and Expert Systems Applications (DEXA'94), Athens, Greece, September, pp. 370-379.

33. Kohavi Ron and Sommerfield Dan (1996). Data Mining using MLC++ A Machine Learning Library in C++. Tools With AI, pages 234–245.

34. Levine, N. (2004). The CrimeStat program: Characteristics, use and audience. Geographical Analysis. Forthcoming.

35. Li W., Han J., and Pei J. (2001). CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules, Proc. 2001 Int. Conf. on Data Mining (ICDM'01), San Jose, CA, Nov..

36. Mannila Heikki, Toivonen Hannu and Verkamo A. (1994). Inkeri: Efficient Algorithms for Discovering Association Rules. KDD Workshop: 181-192.

37. Mapquest.com (2007): http://city.mapquest.com

38. Mayhew, D. J. (2002). The Usability Engineering Lifecycle - A practitioner's handbook for user interface design. San Francisco, California, Morgan Kaufmann Publlishers Inc.

39. Micheline Kamber, Winstone Lara, Gon Wang and Han Jiawei (1997). Generalization and Decision Tree Induction: Efficient Classification in Data Mining. RIDE.

40. Monmonier, M. (1989). Geographic brushing: Enhancing exploratory analysis of the scatterplot matrix. Geographical Analysis, 21:81–4.

41. MSN Virtual Earth(2007). http://maps.live.com

42. Ng R. and Han J. (1994). Efficient and Effective Clustering Method for Spatial Data Mining, Proc. of 1994 Int'l Conf. on Very Large Data Bases (VLDB'94), Santiago, Chile, September, pp. 144-155.

43. Nielsen Jakob, N. N. G. (1993). Usability Engineering. Cambridge, Academic press limited.

44. Piatetsky-Shapiro Gregory(1989). Knowledge Discovery in Real Databases. IJCAI-89 Workshop.

45. Redman, T. (1996). Data Quality for the Information Age. Boston, London, Artech House.

46. Rey, S. J. and Janikas, M. V. (2004). STARS: Space-time analysis of regional systems. Geographical Analysis. forthcoming.

47. Symanzik, J., Cook, D., Lewin-Koh, N., Majure, J. J., and Megretskaia, I. (2000). Linking ArcView and XGobi: Insight behind the front end. Journal of Computational and Graphical Statistics, 9(3):470–490.

48. Takatsuka, M. and Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. Computers and Geosciences, 28:1131–1141.

49. Tayi, G. K. and Ballou D. p. (1998). Examing data quality. Communications of the ACM. 41 (2): 54-57.

50. Unwin, A. (1994). REGARDing geographic data. Computational Statistics, pages 345–354. Physica Verlag, Heidelberg.

51. Wang, R. Y., Strong D. M., (1999). An information quality assessment methodoligy. Proceedings of the International Conference on Information Quality, Cambridge MA.

52. Wise, S., Haining, R., and Ma, J. (2001). Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. International Journal of Geographic Information Science, 15(3):239–254.

53. Yahoo Maps, Driving Directions, and Traffic (2007). http://maps.yahoo.com

54. Yoo J. S. and Shekhar S. (2006). A join less approach for mining spatial collocation patterns. *IEEE Transactions on Knowledge and Data Engineering,* Vol. 18, No. 10, October.