Navigating and Analyzing Geospatial Data Using Open Source Software

Aiman Jawaid Haider
DA-IICT, Gandhinagar
201101058@daiict.ac.in
Supervisor
Dr. Manoj Pandya
BISAG, Gandhinagar.
Dr. Anil Roy (On-Campus Mentor)

***Abstract*-The Potential of GIS technology in analyzing the huge quantum of data is still emerging as a field. The work tries to make an endeavor to contribute to its emergence by creating certain functionalities. The report elucidates how GIS buffer analysis can be done in Open source, which will increase its reach and access. Further the work endeavors to create a generic system, which can be extended further to create any specific application be it aiding individual decisions or decision making at the administrational level.**

***Keywords*- Open Source GIS, Generic Buffer Analysis, Amenity Searcher, Decision Support System, GIS Data Analysis**

I. INTRODUCTION

Information and Communication technology has revolutionized the lives of people ranging from the introduction of social media at individual level to aiding organizations in decision-making. Data creation of the real world activities and their processing has been at the heart of this technology.

Geographic information system has also contributed its part in terms of Navigation systems, Satellite sensing and creation of the maps, finding amenities, etc. Like any other Information and Communication Technology, the heart of GIS is also located in data and its management.

The data that GIS usually captures, stores, manipulates, analyzes and presents is known as Geo-spatial data. Geo-spatial data refers to information describing geographic locations, natural or man-made features and territorial boundaries on the globe. It is represented in various forms, which include points, lines, polygons, 2D, 3D and many higher levels of abstraction. It comprises both directly captured and manipulated data.

From the above, the importance of analyzing geospatial data becomes self-evident having potential enough to improve the lives of people from individual to societal levels. However, most of the geospatial analysis tools are proprietary in nature. This limits the potential of the technology.

The work aims to overcome this limitation, being based on the emerging technique of buffer analysis for the same. Buffer Analysis refers to a technique of data retrieval, which helps in analyzing information around a specific feature of interest. The feature can be in any shape as defined by its geometry. The technique usually consists of creating a buffer around the midpoint/centroid of the geometry, and retrieving data coming inside the buffer radius. A buffer refers to a locus of constant distance from a point.

The following sections would describe the work from the basics of GIS to the creation of a system for generic buffer analysis that can easily be used as a real-life solution. The methodology used was waterfall and is summed up through the sections. Section 2 deals with the background, supporting and complementing technologies. Section 3 deals with System Design and implementation method. Section 4 would include the analysis and testing, Section 5 some specific applications, and finally conclude with Section 6.

II. BACKGROUND

Generally, a GIS consists of the capture, transfer, validate and edit, store and structure, restructure, generalize, transform, query, analyze and present. While the first three phases pertain to the raw data collection and management, then next two phases structure the data to be used further. The next three phases further standardize the data collected from various sources to be worked upon. This collection of data is then queried, analyzed and finally the result is presented. For this purpose, various technologies like a database (e.g. PostGIS, Oracle Spatial) and a map renderer (e.g.GeoServer, ArcGIS, MapGuide) are used among others.

The work in the present context concentrates on the phases after the raw data has been stored. The process then includes the following:

- Restructuring, generalizing and transforming through applications like ArcGIS, GeoServer, MapGuide. The data is taken in various formats. In this work,

ESRI and PostGIS database formats are used as inputs.
- Spatial Database connectivity to analyze the data. The applications for this include Oracle Spatial, PostGIS, Server Spatial, etc. These are also used for querying.
- Presenting the map and related features on the Web Browser. GoogleMaps, OpenLayers, etc. are used to accomplish this task.

For further details please refer to [1].

From the analysis of the above technologies, it was decided to use GeoServer, PostGIS and OpenLayers to accomplish the functionalities of Map rendering and styling, layer switching, information retrieval, finding point from data, measuring data and area, creating buffer and data operations on it.

A. Supporting Technologies

- GeoServer: Aids in publishing of maps, accessing database information attached to the map, converting PostGIS tables into maps and also applying dynamic styling. This is used in the backend.
- OpenLayers: Aids in rendering of maps, measuring distance and areas, displaying buffers/points on map, layer switching, zooming and moving maps on screen. It is used in frontend along with Java (JSP, Java script).
- PostGIS: Aids in the data management of geospatial data with geometry, creation of geometrical buffers and other common queries for data operations. It is used for backend database operations. It is an extension of PostgresSQL.

B. Comparative Analysis

The system created is similar to the following systems. A distinction is created as follows:

- Arc GIS package: It allows for all the above functionalities, but is proprietary in nature.
- Geoserver: It allows for the creation of buffer and some other functionality. However, this can only be done through coding and requires knowledge of the technology.
- Boundless OpenGeo Suite and the like: They allow for plenty of functionalities in Open Source. This system on the other hand has lesser number of functionalities but is in web-environment. They are generally desktop tools.

III. SYSTEM DESIGN

The system has been designed to use the technique of buffer analysis. It presently uses GeoServer 2.6.1 , PostGIS 2.1.4 and OpenLayers 2.0 for technology support apart from web environment for hosting (JavaSript Pages and Java Script). NetBeans IDE 8.0.2 is used for coding. Any other technology as mentioned above can also be used. These technologies give it the open source character. The system has been made user-friendly and is generic in nature i.e. any geospatial data from micro to macro can be used. Further, the system has been constructed such that even while analyzing the data, it can easily be extended from an amenity finder to a decision support system. However, it is presently restricted to point geometry and is for two-dimensional data. Extensions to geometry can be easily done based on similar design. The system implements a number of major and minor functionalities, which are described in the following paragraphs.

Data in the form of Map name is taken as input from the user as the system is generic. It is, however, ensured that the map data is in PostGIS or ESRI format only (by providing a drop down for choosing the map name). This data is present both in the database (PostGIS) and GeoServer, which transforms the map into image. This map image is rendered through OpenLayers. All throughout, the data in PostGIS and GeoServer has been connected.

The data/map inputted can be applied various styling features like color, different legends and labeling with the help of StyleLayerDescriptor. These styling effects are also taken as input from user and dynamically applied through OpenLayers and GeoServer. This data is then rendered onto the screen with layer switching (with the help of OpenLayers) to allow the user to switch between maps. This data/map(s) can be subjected to any functionality described in the following paragraphs. The screen is divided into two parts the display and the workspace. Any change in the workspace reflects on the display area.

Measuring the area and distance on the map is an essential functionality. This is provided with the help of OpenLayers. This allows the user to calculate distance along a series of straight lines while area is measured through creating a polygon of any side.

The connectivity between data and map has been harnessed by two functionalities - Get Information and Finding Points on the map. Get Information takes the help of Geoserver to display the data related to a point on a simple click. Finding a point on the other hand uses the database to fetch details and correspondingly selects its geometry attribute. This geometry attribute is stored in a new table, created to render a new map on the screen with different styling. The data in database is fetched through JSON and Ajax. The map is then rendered through GeoServer and OpenLayers. The styling can alternatively be done with markers of OpenLayers also.

Creating buffer is at the heart of the system. The system allows for the creation of any number of buffers to be created around a single point and for any number of attributes for any

number of maps rendered. A number of buffers on different points can also be created. This is done via taking the essential inputs from the user such as the number of buffers, their respective ranges, color of the buffers, attributes required, etc., Then the database is accordingly queried to create the buffer geometry. The geometry consists of a circle as line joints. This is rendered onto the screen like for finding point in previous paragraph. This geometry is then applied as a circumscriber to the data of interest to be fetched from inside the buffer range (present as attributes in different maps). The data has corresponding latitude, longitude and geometry attributes, which are used to check its presence or absence in the buffer. If the data falls in the buffer range it is placed in another new table/map created specially for the data present in that buffer. The presence of buffer data in table form is then used for undertaking data operations on it and also displaying. The buffer so created is displayed with the help of GeoServer and OpenLayers. Alternatively, the buffers can also be created using the function for the same in GeoServer.

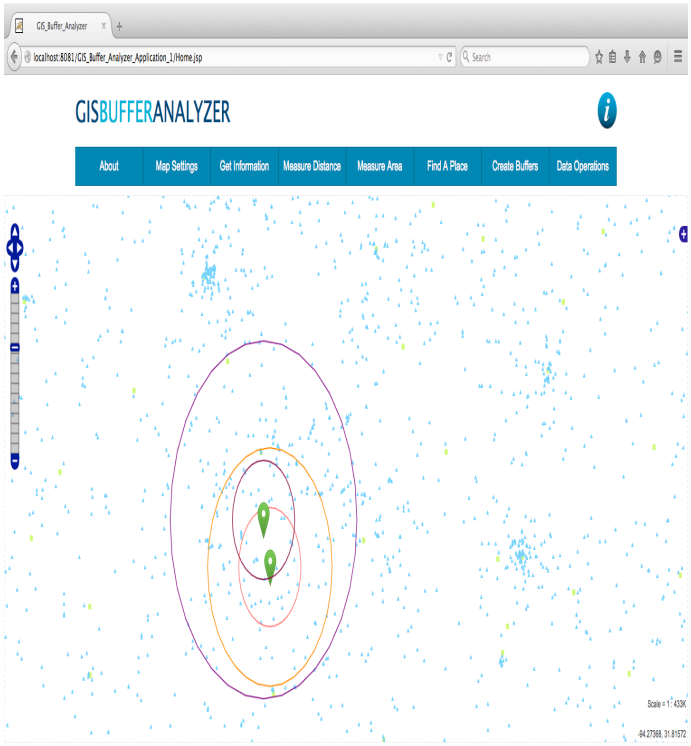


Fig. 1 Example of multipoint multiple buffers created

Analyzing the data created from the above operation is the most important step of the whole work. The data present in the database can be subjected to comparison, intersection between maps (including buffers), union, difference or even simply display as a table and map. This is done through the corresponding queries to be applied on the database. Similar procedure as already described, involving JSON and ajax is used. However, in all the operations mentioned above dynamic queries are created. For displaying the data generated, GeoServer and OpenLayers are used as already stated in the third paragraph of this section.

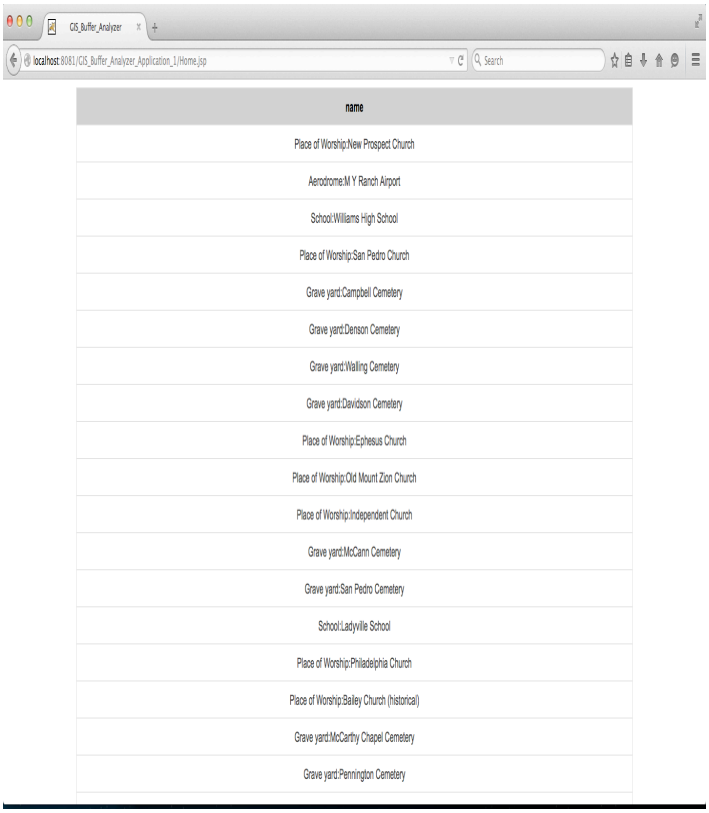


Fig. 2 Example of data fetched from intersection of multiple buffers

IV. ANALYSIS AND TESTING

The system is generic, user-friendly and Open Source. It is, however, recommended that GeoServer and PostGIS (extension of PostgresSQL) apart from JDK be installed in the system. This makes it easy for the user to work with the application in any Operating System and any web browser. However, a number of limitations are also observed.

The data of the map/database being huge sometimes causes time lags. This could be overcome by using distributed computing for faster working.

The system could be well extended to take inputs in other formats like other geometries, dimensions, etc.

For much more independence, it is advisable to create a dynamic manager establishing connectivity between the map renderer server (e.g. GeoServer) and the database (e.g. PostGIS). This would make the system even more user-friendly and raise the level of abstraction from the nitty-gritties of technology.

An extension of the different analyzing techniques would make the system a complete GIS tool which is Open Source and also easy to use.

As the input-output metric involves images, black box testing had been done for each module (functionality) created. Further, bottom-up integration check was done. Every module was tested and then integrated and tested and so on. The results of the testing led to some minor inclusion of checks for the input fields and various other deletion/addition of fields/functionalities so that only certain functionality is used at a time.

V. SOME APPLICATIONS

The generic implementation of all the above-mentioned functionality allows for the use of the system in applications ranging from micro to macro scenarios. For example, the system can be used for locating the nearest dispensaries and even be used to find out places that need to be evacuated during a cyclone. The data generated from the buffers can be manipulated as per the requirement to aid in such wide-ranging scenarios. A mere extension of the system on similar lines can help in creating specific system as per requirement.

The map image allows the user to understand the ground scenario. For example, in case of an outbreak of epidemic a clear picture of the locations of the health facilities can help the administrator in decision-making. In case of the outbreak being near a highly dense network of facilities, the action would not have to be so vigorous as in the case with no facilities around.

Citizens can also use the information around their regions to monitor pollution levels, climate, etc. This would increase the awareness regarding such issues and encourage the citizens to action regarding the same.

A demographic study in a region can also be conducted. The data regarding population, mortality, literacy, etc. can be used to study the trends of demography in a region. This would aid in understanding and in certain cases even addressing the social issues in real-life.

Many such applications and other specific ones can be developed on the basis of this system. Two such are described in the section below:

A. Finding Parking Areas around a Place

Car parking is becoming an amenity of huge importance specially given the growing traffic congestion. The system is made to use map(s) of a certain area containing the parking areas as points. This is then subjected to buffer analysis to find out the parking areas around in the range of some kilometres. The area can be further analyzed with respect to proximity to other amenities also, applying other buffer on it as a centre or even simply visually comparing with the amenities map. The distances along the road can also be compared with the help of the series of measuring lines.

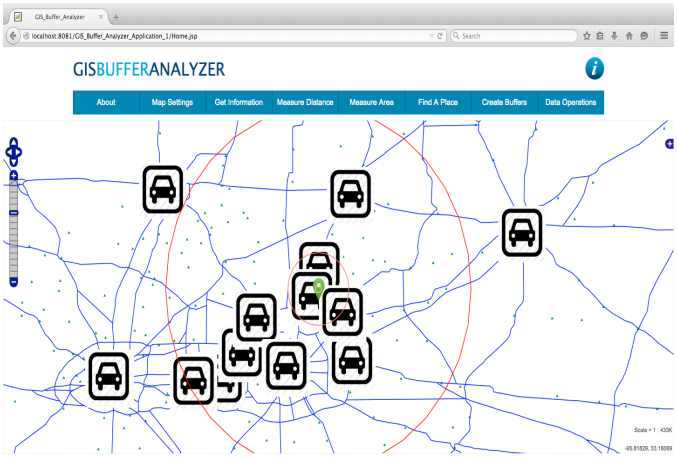


Fig. 3 Creation of Buffer around the place with particular range

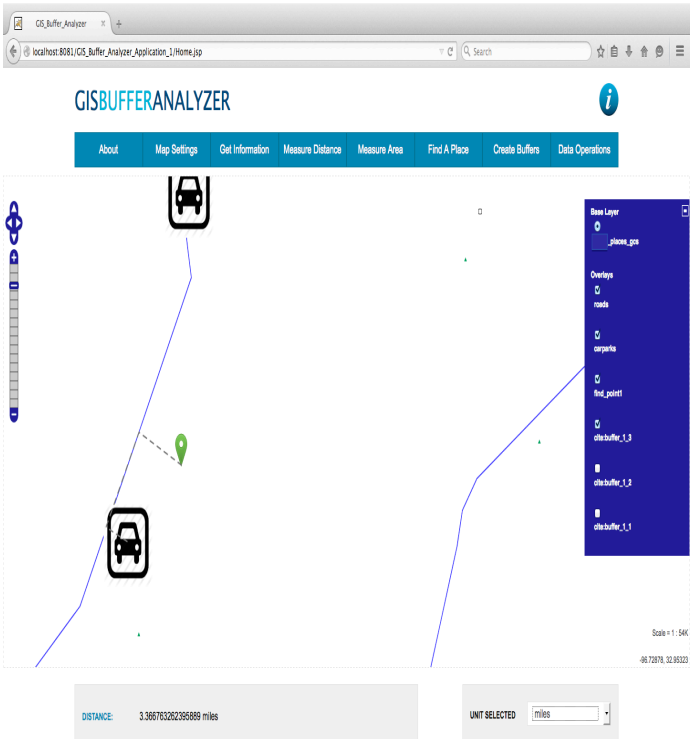


Fig. 4 Finding the distance of a car parking from the specific place

B. Studying the density of Schools and its correspondence with the respective population

The maps containing data for population, schools and territorial boundaries are used. For each headquarter, a buffer is created according to the "standard" distance between headquarters and the school. This allows one to fetch the school data and compare the school densities among the buffers. Clearly, this would bring the areas where creation of such facilities is needed after comparing the hospital to population ratio. Extending the system design further, may

result in a specific decision support system for a particular purpose like this.

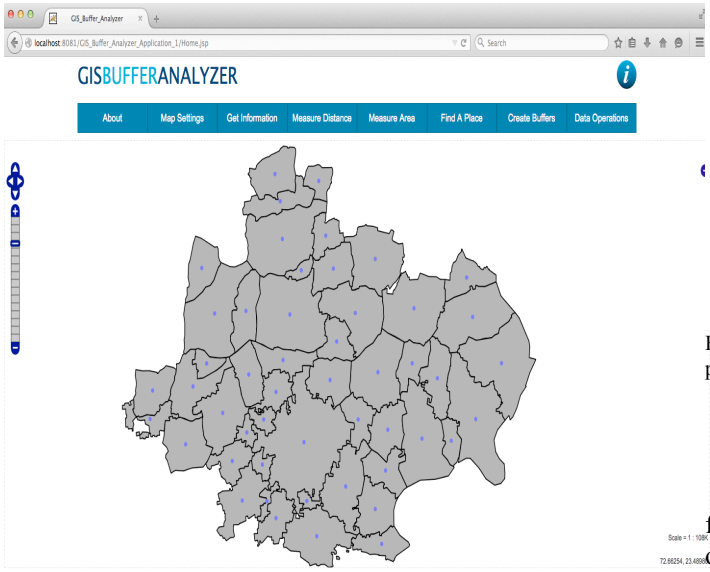


Fig. 5 The map of Mansa Village with the points showing schools

In this particular example, as self evident from the map, one school per village is present. However more analysis is done via the data operations to get a more complete picture.

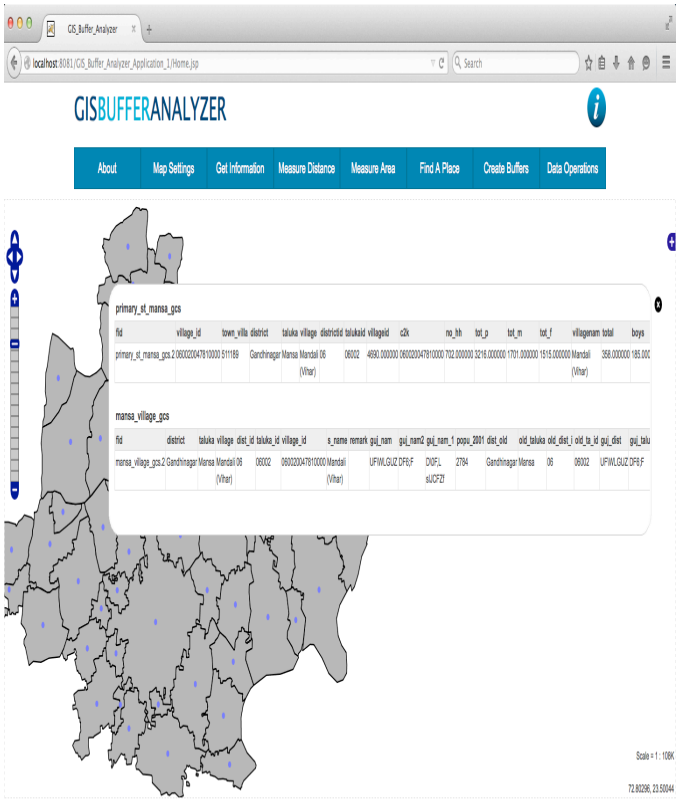
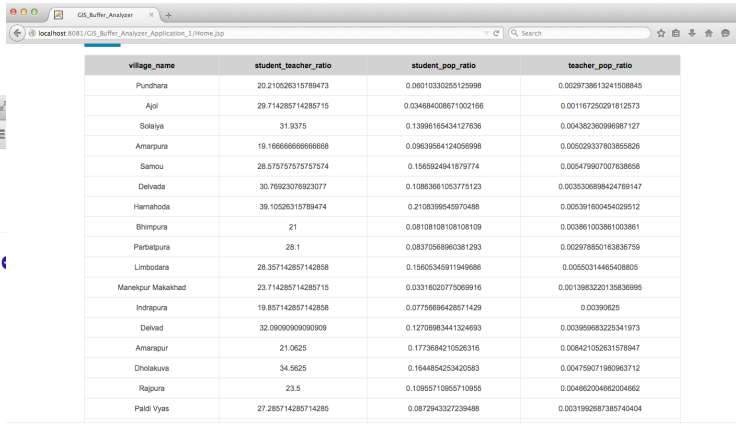


Fig. 6 Finding information



village_name	student_teacher_ratio	student_pop_ratio	teacher_pop_ratio
Pundhara	20.210526315789473	0.06010330255125998	0.002973861324150884
Ajol	29.714285714285715	0.03468400867100216	0.001167250291812573
Solaiya	31.9375	0.13996165434127636	0.004382360996987127
Amarpura	19.166666666666668	0.09639564124056998	0.005029337803855826
Samou	28.575757575757574	0.1565924941879774	0.005479907007638658
Delvada	30.769230769230777	0.10863661053775123	0.003530698424769147
Harnahoda	39.10526315789474	0.2108399545970488	0.005391600454029512
Bhimpura	21	0.08108108108108109	0.003861003861003861
Parbatpura	28.1	0.08370568960381293	0.002978850163836759
Limbodara	28.357142857142858	0.15605345911949686	0.005531446854086805
Manekpur Makakhad	23.714285714285715	0.03316020775069916	0.001398322013583695
Indrapura	19.85714285714285	0.07756696428571429	0.00390625
Delvad	32.09090909090909	0.12706983441324693	0.003959683225341973
Amarapur	21.0625	0.17736842105263316	0.008421052631578947
Dholakuva	34.5625	0.1644854253420583	0.004759071980963712
Rajpura	23.5	0.10955710955710955	0.004662004662004662
Paldi Vyas	27.285714285714285	0.08729433272394888	0.003199268738574040

Fig. 7 Showing the student teacher ratio per village along with student population and teacher population ratios

VI. CONCLUSION

GIS as a technology has huge potential. This can be furthered extended and made more accessible by giving it an open source base and making it more user-friendly. By increasing the reach of this technology, it is bound to bring in more transparency in the working of institutions while also aid the creation of smart systems like earthquake alarms, etc. This apart its help the aid it can give to individuals is also worth mentioning. The proliferation of the digital world is helpful in giving this path-breaking technology a chance to perform. The development of such technologies in Open source and simplifying the technology to the common people would indeed increase the revolutionary power of the technology.

ACKNOWLEDGEMENT

The author wishes to acknowledge Dr. Manoj Pandya, Vyoma Bhatt, Hitesh Bodar and Dr. Anil Roy for their valuable support and inputs in the work. Further, the author wishes to acknowledge the support given by T.P.Singh (Director, BISAG), BISAG, Gandhinagar and DA-IICT, Gandhinagar.

REFERENCE

[1] D.J. Macguire and J. Dangermond, "*The functionality of GIS*", Geographic Information System,1 ed, WILEY, New York: 1991, vol. 1, pp-319-35

[2] S Brajesh, "*Buffer Analysis and Modified K-Means Clustering for Geo-Spatial Data on Gujarat Schools*", International Journal of Scientific & Engineering Research, Volume 5, Issue 6, June-2014

[3] Ramakrishnan and Gehrke, "*Database Management Systems*", 3 ed. New York: McGraw-Hill,2003

[4] C.Horstmann, "*BIG JAVA", 3 ed.* New Delhi: WILEY-INDIA,2002

[5] The ESRI Website [online]. Available: http://www.esri.com/

[6] [ESR98] ESRI, 1998, "*ESRI Shapefile Technical Description*". Available: http://www.esri. com/library/whitepapers/pdfs/shapefile.pdf.

[7] B. YoungBlood, *GeoServer Beginner's Guide,* Birmingham: Packt, 2013

[8] WorldWideWeb Consortium,"JQuery Tutorial"[online], Available: http://www.w3schools.com/jquery/default.asp

[9] Ellie Quigley, *"JavaScript by Example"*, 2 ed, NJ: Pearson Education, 2011

[10] WorldWideWeb Consortium,"JSON Tutorial" [online], Available: http://www.w3schools.com/json/default.asp

[11] Antonio Santiago Perez, *"OpenLayers Cookbook",* 1 ed, Birmingham: Packt, 2012

[12] Paul Ramsey, *"Spatial Database Tips and Tricks Presentation",* Free Open Source Software for Geospatial Workshop, Sydney, 2009, Available: http://workshops.boundlessgeo.com/postgis-spatialdbtips/

[13] IIRS, "Geographical Information System" [online], Available: http://elearning.iirs.gov.in/courses/paper_05/default.htm