

Visual Analysis of Spatial Metadata

Savita Kumari Sheoran and Vinti Parmar

Abstract—Spatial data gathered from various data sources contains huge amount of metadata. Large amount of metadata creates problem for the user to get insight into the required data for any specific application. The problem of selecting required data can be solved by applying data query and spatial query in GIS tool. This paper is attempted to select, visualize and analyze spatial metadata by performing query operations followed by visualizing techniques and plotting graphs to find the densely populated areas in Gurugram District. Spatial data of Gurugram District having large set of metadata information is added to QGIS tool. Query operations followed by dense pixel display visualization technique is used in GIS for better understanding of result data set. The result set having Gurgugram City, Farukhnagar, Sohna, Pataudi and Manesar are further analyzed by plotting graphs.

Index Terms—Data exploration, geographic data, ISO/TC211 standards, spatial metadata, visualization techniques.

I. INTRODUCTION

The development in technologies like big data analytics, cloud computing, artificial intelligence, sensors, wireless networks, high growth in internet users and societal transformation results in rapid growth of data generation as everything on internet is recorded. Each activity performed on internet produces data. With this advancement of technology, spatial data play an important role in day today activities. This vast amount of data is persistently growing, providing consumers an unendingly growing choice of spatial datasets. Spatial data are broadly used by the public sector, private sector and common people for decision making, key planning, risk analysis, and route finding. Geographic data comprise mainly geospatial data. The importance of digital information is growing in business planning, commerce, manufacturing industries, healthcare, agronomics, financial affairs, aerology, experimental research, astrograph, shipping and strengthening of society. Geospatial data involve data related to location which is valuable for designing and configuring of enterprise data stores. Large investments are made in compiling, governing and dispersing information but most of the part is assigned to spatial component. The production and storage of spatial data are tedious and expensive process [1]. Spatial data generation from spatial raw data requires experience,

dexterous and advanced skills. The procedure of exploring new and valuable patterns from vast spatial dataset is known as spatial data mining. Since conventional data handling technique are unfit to handle spatial data, so concept of spatial database and spatial data warehouse came into existence. SOLAP (Spatial On-line Analytical Processing) tools provide high level of data communication to users for representing spatial data. Basically, metadata facilitates with necessary information and meaningful interpretation of sensor data, instrument status and functioning of the observatory [2]. Spatial data can be handled in different views to present information at different levels so that the outcomes can be analyzed in the form of maps, tables and charts etc. [3]. Spatial data is extensively received from various sources and metadata is used to give information about spatial dataset. Metadata is represented and transformed in electronic form in digital archives such as metadata information system [4]. Following issues are related to spatial data:

- 1) Entry of spatial data and metadata.
- 2) Access to various data sources.
- 3) Information integration from various sources.
- 4) Data selection according to user needs.

In this paper we endeavor on identifying of required data from large data sets by data selection and analyzing it using visualization techniques in QGIS Software. Here we use QGIS as it is an Open source GIS program and have user friendly graphic user interface for spatial metadata selection and visual analysis. The main aim of the paper is to discover and understand useful data from large data sets. For this we represent metadata with the following mathematical model:

M represents dataset of all metadata elements, and $M_1, M_2, M_3, \dots, M_N$ are the metadata elements provided with any spatial data. This can be represented using below equations:

$$M = \{M_1, M_2, M_3, \dots, M_N\}.$$

Metadata data set would be taken from data sources $D_1, D_2, D_3, \dots, D_N$.

The required dataset of metadata element IM is,

$$IM = \{M_1, M_2, M_3, \dots, M_i; i < N\}.$$

Thus, $IM \subseteq M$.

Above mathematics signifies that we can select required metadata IM from available data M by applying query operations on M dataset. The result set further can be better analyzed by plotting graphs.

II. METADATA ANALYSIS FOR DATA SELECTION

Geospatial Metadata explains the lineaments of geographic data that are helpful in adhering value to

Manuscript received January 9, 2020; revised April 1, 2020. This work is accomplished to meet the requirement of Ph.D. (Computer Science) degree at Indira Gandhi University Meerpur, Rewari, India.

Savita Kumari Sheoran and Vinti Parmar are with Department of Computer Science and Engineering, Indira Gandhi University, Meerpur, Rewari, India (e-mail: savita.sheoran@igu.ac.in, vintiparmar1487@gmail.com).

geographic dataset. Metadata are data sets that provide important information about other data [5]. It basically belongs to entities that have some spatial extent. Geospatial metadata can be defined as the data that is associated with some location on the plane of the globe. Geospatial metadata is usually required to store geospatial data sets and resources including mapping applications, data models and web based services. There is a big challenge in the existing management and application of geographic data that user still realize that the quality information provided could not able to meet their needs. It could not able to describe the datasets to choose for their required applications. Different organizations follow different standards for documenting data quality information and Metadata. ISO/TC 211 standards are widely used to represent the digital data of geographic earth. These standards specify location based services, temporal schema, imagery standards, reference model, spatial schema [6]. It represents and changes such data in digital form between various users, data producers, various systems and places. There are different metadata standards that are utilized to explain a product to the users. Different organizations follow different standards for documenting data quality information and metadata. ISO/TC 211 standards are widely used to represent the digital data of geographic earth. The ISO/TC211 comprise of 55 national bodies. The standard that comes under ISO/TC 211 is named as ISO19100 family. The ISO 19115 “Metadata” is apparently the best recognized standard of the ISO19100 family. These standards defines reference model, location based services, spatial and temporal schema. It can be applied on digital data and represent data in various forms such as maps, textual documents, charts as well as non geographic data. In data selection activity we try to choose the most appropriate data for a particular application. Data selection depends upon the complexity of geographic data [7]. The two important parameters for spatial data selection are:

1. Correct interpretation of geographic data at a glance.
2. Comparison of various geographic data sources.

Both of the above mentioned activity is very difficult as data may not be freely available every time and geographic data sources differ in terms of scale, reference system, themes etc. So to overcome these difficulties metadata is used.

III. APPROACH FOR VISUAL METADATA ANALYSIS

The Spatial data mining techniques used so far in different fields are incapable in representing the complete metadata descriptions of the geospatial dataset [8]. Earlier different data mining techniques, statistical techniques, geographical and cartographic techniques are used for retrieval of spatial data but access to spatial data quality information were problematic [9], so for better understanding and efficient use of spatial metadata we will use visual data exploration techniques to explore the geographical metadata, where a visualization will enhance communication between the user and the computer. The visualization techniques are classified into dense pixel display, iconic display, standard 2D/3D display, and interaction and distortion techniques [10].

A. Standard 2D/3D Display

Finds interesting transformations of multi-dimensional data sets, data item is presented as a polygonal line, intersecting each of the axes.

B. Dense Pixel Display

Mapping between each dimension value to a shaded pixel and group the pixels belonging to each dimension into adjacent areas, use different arrangements for different purposes.

C. Iconic Display

Mapping between the attribute values of a multi-dimensional data item to the features of an icon, Icons can randomly used. They may be small faces, star icons, stick figure icons, needle icons, color icons.

D. Interactive and Distortion Techniques

Allow users to communicate with the visualizations by providing interactive Zooming and Filtering. See Fig. 1:

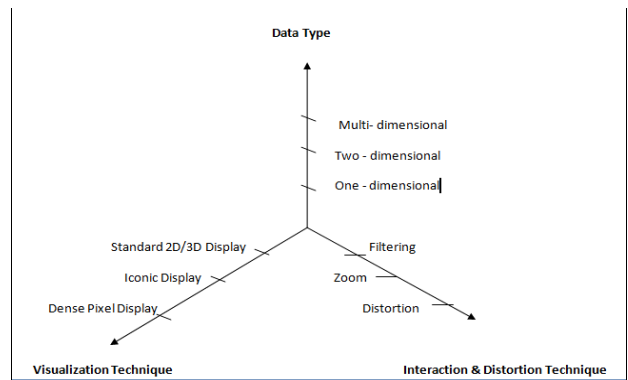


Fig. 1. Classification of information visualization techniques.

In the following section of the paper we describe an approach to analyze metadata using the Dense Pixel Display visualization technique for visual analysis of metadata.

E. Illustrative Example

We illustrate an example to study the population distribution in different villages and towns of Gurugram district using QGIS Software.

Metadata dataset. In this study, spatial data of Gurugram District in shapefile format will be added to the vector layer of QGIS tool. The spatial data consist of Boundary Id, Area and Name of all the rural and urban regions. Large set of metadata containing information about District code, District Name, total population census 2011, literacy of total population etc. will be added to this Spatial data to get insight about various regions. We will find the highly populated areas from large metadata set by applying query operations and the result set will be visualized and analyzed for better understanding of the data. Performing query on the data in the attribute table of QGIS software will give result in textual form which is quite difficult to understand. So user needs to perform visual analysis in the repository to satisfy his requirements. For this below mentioned five iterative phases are used:

- (1) **Data Creation:** In this phase vector data is added to the QGIS layer. User added spatial data of Gurugram District (Fig. 2) in vector form to QGIS which contain information about District code, District Name, total

Population census 2011, literacy of total population etc. but without applying any visualization technique data is difficult to understand. All town and villages are shown with same color. No information is given about attribute on layered data.

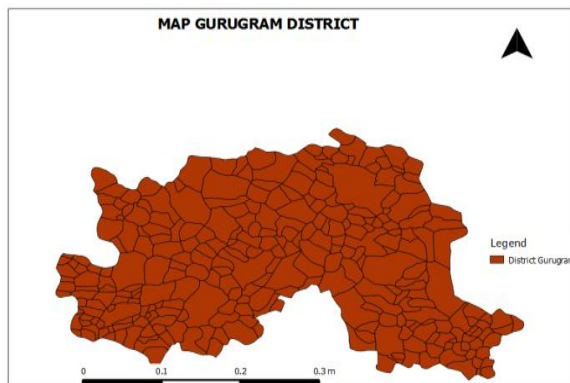


Fig. 2. Map Gurugram district.

- (2) **Information Extraction:** Information about the spatial data is extracted using the attribute table and properties of the data. User interacts with the software and open metadata (Fig. 3) in attribute table which is difficult to understand and compare. Large data sets having various attributes like District code, District Name, Tehsil Name, total Population, literacy of total population etc are provided in attribute table.

EB-0618_Level	EB-0618_Name	EB-0618_TRU	EB-0618_No_HH	EB-0618_TOT_P	EB-0618_TOT_M
VILLAGE	Mangwaki(11)	Rural	128	675	342
VILLAGE	Haliaki(13)	Rural	156	810	425
VILLAGE	Nurgarh(12)	Rural	384	1822	938
SUB-DISTRICT	Pataudi	Urban	7454	41324	21745
SUB-DISTRICT	Pataudi	Rural	15122	78688	41060
SUB-DISTRICT	Pataudi	Total	22576	120012	62805
DISTRICT	Gurgaon	Urban	237750	1042253	565228
DISTRICT	Gurgaon	Rural	88678	472179	251462
DISTRICT	Gurgaon	Total	326428	1514432	816690
VILLAGE	Bahmanwas(10)	Rural	73	339	182
VILLAGE	Ghudana(9)	Rural	416	2075	1074
VILLAGE	Husainka(15)	Rural	202	1079	558
VILLAGE	Rajpura(8)	Rural	212	999	539
VILLAGE	Muzaphara Alia...	Rural	311	1464	781
VILLAGE	Sherpur(16)	Rural	422	2207	1172
VILLAGE	Chhilarki(14)	Rural	157	772	416
VILLAGE	Jasat(17)	Rural	212	1177	556
VILLAGE	Brjipura(18)	Rural	113	554	307
VILLAGE	Barheri Rehnwa...	Rural	193	939	495
VILLAGE	Dadawas(20)	Rural	163	824	418
VILLAGE	Haqdarapur(19)	Rural	240	1334	710
VILLAGE	Shahpur Jat(263)	Rural	174	988	520
VILLAGE	Inchhapuri(264)	Rural	368	1831	968
VILLAGE	Milakpur(3)	Rural	178	912	484
VILLAGE	Mirzapur(4)	Rural	269	1355	703
VILLAGE	Dewlawas(5)	Rural	82	391	201
VILLAGE	Chhawani(2)	Rural	159	808	414
VILLAGE	Ransika(31)	Rural	169	973	497
VILLAGE	Khur(32)	Rural	888	4033	2556

Fig. 3. Large metadata set in attribute table.

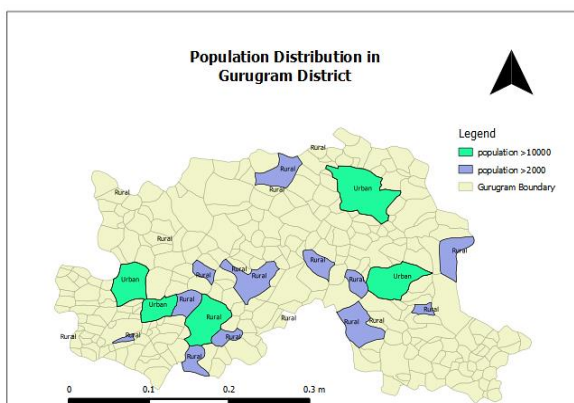


Fig. 4. Population distribution using dense pixel display visualization technique.

- (3) **Visualization:** We apply dense pixel display visualization technique on the spatial data to gain insight about population distribution (Fig. 4) in various rural and urban areas. This data is categorized into different ranges. Regions which are in same range of population distribution are shown with similar pixel colour in the data. Now we can easily analyse that urban areas have high population compared to rural areas.
- (4) **Query processing:** In this metadata is selected from large data set by applying query on the data. We apply query to find highly populated urban areas. Regions having population greater than 10000 are filtered and results are shown below in Fig. 5.

BOUNDARY_ID	AREA	Unique_ID	Name	EB-0618_TOT_P
4	3523300	NA	Farukhnagar	25201
47	86	0.000	NA	225460
130	194	0.000	NA	28811
223	238	0.000	NA	22576
281	298	0.000	NA	24574

Fig. 5. Results after applying query on attribute table.

Now data can be analyzed that five regions Gurugram City, Farukhnagar, Sohna, Pataudi and Manesar have population greater than 10000. These results can be visualized on the added vector layer using visualization techniques with graphic interaction tools. For better data exploration we combine the automatic visualization techniques with graphic interaction tools to find hidden relationships in the large data sets. So for better results examination we combine visualization technique with graphic interaction tools in the next phase.

- (5) **Metadata analysis using graphs:** The selected metadata is visualized by plotting graphs against boundary id and population. Scattergraph is generated to results of query processor. Now User can easily visualize that there are five regions having population greater than 10000 as shown in Fig. 6.

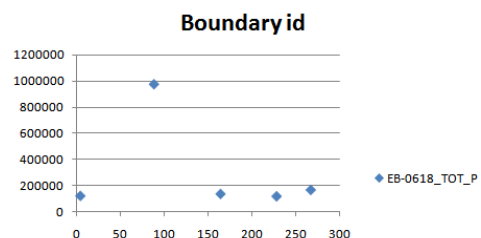


Fig. 6. Graphical representation of highly populated areas.

IV. DISCUSSIONS

In this paper we analyzed spatial metadata in real case study using QGIS software. Spatial data of Gurugram District containing various information like Total

Population, female population, male population, literacy of total population etc. is added to the vector layer. The population distribution in different villages and towns of Gurugram district is studied. Performing open attribute table operation on the spatial data shows data containing large sets of data in textual form which is difficult to analyze. So we used dense pixel display visualization technique for visualizing data which is categorized into different ranges of population distribution. Regions which are in same range of population distribution are shown with similar pixel colour in the data. Then we find densely populated area by processing query to get the regions having population greater than 10000, the result set contain Gurugram City, Farukhnagar, Sohna, Pataudi and Manesar. Further result set is analyzed by plotting graphs against boundary id and population. Scattergraph is generated to the results of query processor.

V. CONCLUSIONS

In this paper we described how to select and analyze metadata using the Dense Pixel Display visualization technique for visual analysis of metadata. User added Gurugram district spatial data in vector form to the QGIS layer which contains large sets of data having information of Total Population, female population, male population, literacy of total population of Gurugram district. The visual approach gives precise view of the data and hence user became capable to analyze which is the required and relevant data from the available large set of data. The result set selected and analyzed by this tool will be used in performing multi-criteria analysis for Landfill site selection in Gurugram District. In similar way required and relevant metadata can be selected and analyzed for different applications like experimental research, transportation etc. depending upon the metadata provided with spatial data to fulfill different requirements of users. It gives better clarity about the results.

CONFLICT OF INTEREST

The authors declare no conflict of interest. No external funding was acquired to complete this research.

AUTHOR CONTRIBUTIONS

Vinti Parmar has conducted this research under the supervision of Dr. Savita Kumari Sheoran.

ACKNOWLEDGMENT

The authors duly acknowledge the Municipal Council Gurugram for providing data and Department of Computer Science & Engineering, Indira Gandhi University Meerpur,

Rewari for providing facilities for this research.

REFERENCES

- [1] K. Hamad and C. Quiroga, "Geovisualization of archived ITS data-case studies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 104-112, Jan 2016.
- [2] C. Kessler, M. Raubal, and C. Wosniok, "Semantic rules for context-aware geographical information retrieval," *EuroSSC*, pp. 77-92, Sep 2009.
- [3] X. Meng, Y. Wang, and Y. Wu, "Modeling geospatial sensor knowledge under a semantic sensor web environment," in *Proc. IEEE 17th International Conference on Computational Science and Engineering*, pp. 1090-1095, Dec 2014.
- [4] J. Zhang, J. Ma, Y. Zhao, B. Xu, L. Sun, J. Chen, and X. Yuan, "Visual analysis of public utility service problems in a metropolis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1843-1852, Dec. 2014.
- [5] H. Singh and D. Agrawal, "A meta-analysis on content based image retrieval system," in *Proc. International Conference on Emerging Technological Trends (ICETT)*, pp. 1-6, Oct 2016.
- [6] I. Nevat, G. W. Peters, F. Septier, and T. Matsui, "Estimation of spatially correlated random fields in heterogeneous wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2597-2609, May 15, 2015.
- [7] X. Pons and J. Maso, "A comprehensive open package format for preservation and distribution of geospatial data and metadata," *Computers & Geosciences*, vol. 97, pp. 89-97, Dec. 2016.
- [8] D. A. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1-8, Jan.-March 2002.
- [9] A. Bharati, D. Moreira, J. Brogan, P. Hale, K. W. Boyer, P. J. Flynn, A. Rocha, and W. J. Scheirer, "Beyond pixels: Image provenance analysis leveraging metadata," in *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1692-1702, 2019.
- [10] T. Zou, W. Li, P. Liu, X. Su, H. Huang, Y. Han, and X. Guo, "An overview of geospatial information visualization," in *Proc. IEEE International Conference on Progress in Informatics and Computing (PIC)*, pp. 250-254, 2018.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Savita Kumari Sheoran is presently an associate professor in Department of Computer Science & Engineering, Indira Gandhi University Meerpur, Rewari, India. She had received her Ph.D. in computer science from Banasthali Vidhyapeeth (Rajasthan), India and possesses more than fifteen years of experience in teaching and research supervision in various reputed Institutes and Universities in India and abroad.

Dr. Sheoran has authored 6 books / chapter and published 64 research papers in international and national journals conferences / seminars / workshops. She is an active researcher having interest in research domains of mobile computing, social media computing, big data analytics and crime prediction.



Vinti Parmar is an assistant professor in GCW Gurawara, Rewari, India. She has been pursuing for her doctoral degree at Department of Computer Science and Engineering, I.G. University, Meerpur, Rewari on information quality management in geospatial metadata. Her research interests include data mining, text mining, and big data analysis.