IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-8 E-ISSN: 2347-2693

Design and Development of tool for assessing OpenStreetMap Completeness

Sonali Arora[1*], Sukhjit Singh Sehra[2], Sumeet Kaur Sehra[3]

[1*]Dept. of Computer Science & Engineering, GNDEC, PTU, Ludhiana, India
[2] Dept. of Computer Science & Engineering, GNDEC, PTU, Ludhiana, India
[3] Dept. of Computer Science & Engineering, GNDEC, PTU, Ludhiana, India

[*]Corresponding Author: sonali.arora4529@gmail.com

Available online at: www.ijcseonline.org

Received: 02/Jul/2017, Revised: 14/Jul/2017, Accepted: 10/Aug/2017, Published: 30/Aug/2017

***Abstract*—** OpenStreetMap (OSM) is a cumulative effort to create a free presentable map of the world that can be accessed by anyone. OSM is one the most prevalent instance of Volunteered Geographic Information (VGI). Since OSM is generating large amount of spatial data that has been contributed by users with the different level of mapping experiences and different backgrounds, hence the quality of OSM data can vary strongly. For this different studies have been done in which different aspects have been investigated. In most of the studies, ground truth reference datasets have been used for comparison of the data which is called the extrinsic analysis. But extrinsic analysis is not always possible because of lack of availability of ground truth reference datasets. Hence, intrinsic analysis can serve as prominent basis for making the approximate statements on the quality of OSM. The investigation analyses the existing intrinsic frameworks and its limitations and then proposed the new six quality parameters for assessing the completeness of the OSM data effectively. A framework has been developed on the basis of proposed parameters. The results obtained from execution helps in doing the statistical analysis and interpretation by providing the visualizations in the form of bar charts, graphs, tables and maps to assess the completeness of data without the help of any ground truth reference datasets. This enables arbitrarily OSM completeness assessment for any part of the world.

***Keywords*—** OpenStreetMap; Data Quality; Volunteered Geographic Ingormation ; Completeness

I. INTRODUCTION

OpenStreetMap(OSM) is a computerized map database of the world worked through crowdsourced Volunteered Geographic Information (VGI) [1]. The information from OSM is freely accessible for visualization, inquiry, download, and editing by open licenses. OSM works in a manner like Wikipedia where practically all components are open for editing by any individual from the user community [2]. The OSM gathers a huge measure of geo information and makes it freely accessible to the world. The users are encouraged to contribute to the information and provide the dynamic content with the evolution of Web 2.0. OpenStreetMap is using crowdsourcing and following the practice of allowing the users to do the mapping activity as per theirdd knowledge level. Since there are a variety of users contributing to OSM with a different level of experience, it is very important to assess the quality of OSM.

Most of the research studies have focused on assessing the quality using the government datasets. But the major limitations in these were the limited availability of these datasets and licensing restrictions. Hence, it is necessary to create a suitable alternative. This investigation focused on introducing the quality indicators and then designing the tool incorporating them that can help in giving the approximate comments on the quality of OSM. Absolute comments can only be made in the availability of government datasets. The investigation analysis the OSM data obtained from calling the OSM API and manipulates the data to make the statistical interpretations so as to assess the completeness of the selected region. The tool has web interface that has been designed using the Bottle API and python scripts that manipulates the data obtained from OSM server. The results are obtained in the form of bar charts, graphs, tables, maps that makes the statistical interpretation easy [1-25].

A. OpenStreetMap

OpenStreetMap is a mapping project involving the volunteer contribution to create the map of the world. Since there are limitations on using the various available map of world, OSM has emerged as an alternative to create a free editable map that can be used by anyone. The advent of inexpensive portable navigation devices also has contributed to its growth. Steve Coast found the OSM after getting the inspiration from the Wikipedia success. Anybody can contribute to OSM by gathering the information from the GPS devices, manual surveys, digital camera or any other source. Also the aerial photography from Yahoo and the government datasets also has contributed a lot in enriching

the data content of OSM. The data is added to OSM under Open Database License [22]. There are various software available for editing the data in maps such as JOSM , Potlach, GNOME etc. OSM has also assisted many times in case of natural disasters, for example, the OSM team worked hard to map the parts of Haiti using satellite imagery during HAITI earthquake in 2010 to help the assisting team with mapping the refugee camps.

OSM data mainly comprises of four elements which are nodes, ways, relations and tags. Nodes are basically the point features having coordinates consisting of longitude and latitude. Ways are ordered list of nodes, called the linear features. Relations are utilized to combine the geographically associated objects together. Tags are key match pair that is utilized to portray the attributes of an object. OSM's main database is PostgreSQL which contains one table for each element. There are several other formats also available. The main dump file of OSM is planet.osm which contains all the data of OSM since its creation. Many famous companies have now incorporated OSM in their project such as Apple, Flickr, Craiglist, MapQuest etc.

B. Completeness of OpenStreetMap

Quality is the most important aspect of geodata. Completeness is one of the key components of data quality. "Completeness" describes how complete a dataset is. A surplus of data is referred to as "Error of Commission", a lack of data in contrast as "Error of Omission". This completeness is considered in both the aspects i.e., the attribute wise completeness and feature wise completeness. Hence, it is very important to assess the OSM completeness for confidently using OSM in further map production or exchange of information [3]. Assessment of OSM can be done in two ways- one is extrinsic analysis that involves the comparison of OSM data with the ground truth reference datasets and the other is intrinsic analysis that involves proposing new quality indicators and doing the statistical interpretation using them. The extrinsic analysis is difficult to carry out due to licensing restriction and unavailability of ground truth reference datasets, Hence the intrinsic approach allow us to make the approximate comments on the completeness of OSM for any part of the world.

The remainder of this paper is organized as follows. Section 2 summarizes related scientific work done on OSM data quality. Before leading to methodology used, several indicators has been proposed for doing the quality assessment in Section 3.Section 4 explains the methodology that has been used. Section 5 assesses the results obtained after running the python scripts and hence, manipulating the data. Finally, Section 5 concludes the work by summarizing the results of this investigation and discusses future scope.

II. RELATED WORK

Checking the completeness of crowdsourced data is a significant research topic in field of geodata. Different studies have been done for assessing the completeness of OpenStreetMap data. Mostly the studies involved the extrinsic analysis in which data of OpenStreetMap is compared with ground truth reference dataset. But due the unavailability of these reference dataset, intrinsic analysis approach can help us in commenting on the completeness of OpenStreetMap.. For instance, Haklay [4] demonstrated the goodness of the geographical information and also compared the OpenStreetMap with Ordnance Survey Datasets. The main goal of this research was to fill the gap by analysing the information of OSM and concentrating on the analysis of the quality by comparing it with the datasets of ordnance survey (OS). The analysis mainly focused on England and London when OSM was started in London in August 2004. Only streets and roads were used for the purpose of comparison as they are the main features contributed by volunteers. Positional accuracy and completeness were the two elements that were taken into account. The result of the analysis showed that the information of the OSM could be accurate on average within 6m of the recorded position by the OS and around 80% overlap of the motorway object between the two datasets. The OSM dataset is providing only 24.5% coverage of the total area from the completeness point of view of the selected region used in comparison with the ordinance dataset. Sehra et al. [9] also assessed the OpenStreetMap data quality in which he found that with the gaining popularity of OSM, relative errors and number of absolute errors are on decrease. In his research, it was also found that OpenStreetMap is getting mature and developing and will soon reach the maturity of the geodata from the commercial vendors.

Costa [1] explained about the quality and utility of Volunteered Geographic Information (VGI). The research mainly focused on the OpenStreetMap (OSM) building block applicability which is used for the official spatial databases. The result concluded that the town centres completeness of the OSM data is very high, but its value decreases when going away from the urban centres. The research also demonstrated that the attribute completeness relies on the level of urbanization and the nature of the attribute. Fan et al. [8] presented the framework of OSM building footprint data for the intrinsic quality assessment. For the quality measurement, seven quality indicators were suggested without using any of the reference data. The result found from the experiment shows that the data which was present on building the footprints are mapped with the higher accuracy and the quality in terms of positions, geometries, and semantics which is improving with the time.

Barron et al. [5] described the comprehensive framework for the intrinsic OpenStreetMap quality analysis. The

investigation in his paper contained a framework in which more than 25 indicators and methods were used which allows the assessments quality of OSM which is based on the data history. The approximate statements on the OSM data quality could be made without using the reference data sets. To fulfil this purpose, existing methods were taken and developed further to integrate them in the extensible open source framework which enabled the arbitrarily repeatable intrinsic OSM quality analysis for any region of the world.

Mooney and Corcoran [6] analysed the interaction and co-editing patterns of the OpenStreetMap Contributors. The study discussed the results of the social network which were based on the analysis of the seven major cities in OpenStreetMap (OSM) in order to understand the quantitative evidence of the collaboration and interaction among the OSM members in these mentioned areas. Girres and Touya [14] assessed the data quality of the OpenStreetMap by extending the work of Haklay to France and hence considered a larger dataset and used new quality measures. Zhao et al. [23] analysed the evolution of OSM road networks in the city of China, Beijing that focused on the evolution from the centrality analysis, general analysis, topological analysis, and geometric analysis. The general analysis indicated the overall growth pattern and their mapping contributions. The geometric analysis suggested the mapping intensity which exhibits from the heavy-tailed pattern, mapping direction move from outskirts to downtown, and the mapping behaviour of the road networks. Topological analysis indicated the OSM road networks which resemble the growth of real road networks and the evolution process of densification and exploration.

III. DEFINING THE INDICATORS FOR QUALITY ASSESSMENT

There are six quality indicators that have been taken into account for assessing the completeness.
These are:

A. Particular place and way details in brief
B. Count number of building, atm, ways and places
C. Attribute Completeness
D. Assessing the version history of nodes
E. Count number of nodes, ways and buildings created in last five years
F. Development in the no. of buildings and areas like town, villages and others

Each indicator is discussed below:

A. Particular place and way details in brief

Points-of-Interest are represented by nodes, ways and relations in OSM geographically. Restaurants, sights, bus stops are different examples of various points-of-interest and are very important from the point of complete information availability about them. OpenStreetMap contains the information about these POIs in the form of key-value pair. The tool will represent the map containing the brief information about the various places, ways enclosed in that region. Hence, the map containing the brief detail of the POIs attributes allows statements on the relative attribute completeness of that region.

B. Count number of building, atm, ways and places

This indicator will count the number of buildings, ATMs, ways and places. If possible, it could then be compared to Google's map or some other map with the listed number of buildings, ATMs, ways, places and tolls of the same selected region to do some comments on the feature completeness. As Google map data is kept under professional supervision and is updated continuously, significant comments of that region completeness feature wise can be made by doing the comparison with the Google data. In fact, in future, if we have the availability of ground truth reference dataset, this count could be compared with that for checking the feature completeness.

C. Attribute Completeness

Goodchild [24] expresses Completeness as the degree to which the attribute descriptions of a feature are captured. It is 100 percent if all the relevant attributes of a feature are available corresponding to a given capture specification. OpenStreetMap provides us lots of tags to be added while plotting a point, way or relation. These tags give the significant information about the corresponding data type respectively. If all the attributes corresponding to that data type are present, then completeness is enhanced.

D. Assessing the version history of nodes

The history data of OSM can be used to make some comments on completeness of OSM data. A new node version is created whenever a node is updated. If the number of tags increases version wise, then comments on its attribute completeness enhancement can be made.

E. Count number of nodes, ways and building created in last five years

The evolution of a region over a specific time can help us in making the comments on completeness by providing a vision into the development and quality of that region. Hence, bar chart will allow us to interpret the qualitative and quantitative development of these regions. If we have very small yearly increase in number of that region, then it can be stated as almost completed. This thought is supported by fairly high number of contributors who are participating actively to develop the regions which are less mapped. This is the reason that they simply are not doing mapping in that region because of near completion and hence moved to other regions. Hence, this is only an approximation method for

quality parameter completeness. Exact comments can only be done in case we have ground truth reference dataset. But we can benefit from this indicator for making the approximate statements on OSM quality completeness.

F. Development in the no. of buildings and areas like town, villages and others

More is the evolution of OSM features over a specific period of time can be associated to more development and quality of an arbitrarily chosen area within OSM. For example, the increasing number of modified or created points, lines and polygons per year gives a first general and more diverse impression of an area. However, rating these edits in terms of good or bad quality heavily depends on the individual case of the experience of contributor. The quantitative calculations mentioned above allow only limited statements on data quality in some cases.

IV. METHODOLOGY USED

The main aim of the research work is the assessment of completeness of OSM data. The OpenStreetMap is used in this research to access the OSM data and further that data is used for checking the completeness. The OsmApi is used as a sole input for the analyses. It is RESTful api and takes the requests in the form of HTTP PUT, GET, POST, DELETE etc. There is only the restriction on the amount of data that can be exported from this API but apart from this, any region can be checked for the completeness. Also, as there is no deployment of reference dataset for evaluation of quality, hence, intrinsic analysis can be carried out for any region without any reference dataset. But exact statements on completeness cannot be made; it can only serve the relative statements on completeness. The framework is designed to implement the quality indicators. The methodology adopted for the complete design and implementation of the tools can be clearly understood from the flow chart given below:

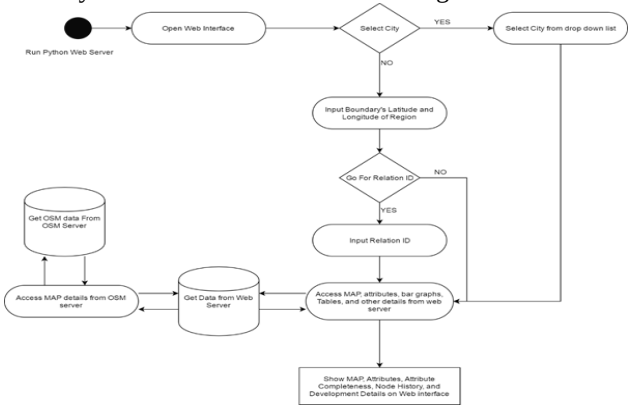

Figure 1. OSM Architecture Framework

Below is the stepwise explanation of flow chart :

a. A tool is created for checking the completeness of data using OSM API.
b. The nodes, building, and ways are counted in a particular region.
c. The region is checked for development that how the area is developed in last five years.
d. The bar graphs drawn using matplotlib show the buildings, nodes and ways created in last five years.
e. The node history is used for checking the average node mean versus version chart to understand the pattern of attribute completeness of nodes version wise.
f. A web interface is developed using Python bottle API which provides a better interface to user through which the user can input the BBOX values and after that it will show the result which is sent to web interface from the python bottle server
g. The result on web interface shows the bar chart, tables, and map of resultant data by using the python server.

V. RESULTS AND DISCUSSION

After the long implementation, the tool can be acknowledged with completed functionalities. The OSM gave many results which are categorized by their types. There are several techniques and algorithms used to check OSM data and by analysing them, we get the results and finally interpretations were made from that. A small region of New Delhi has been considered as a case under study. Results which were achieved after performing the analysis are obtained in these ways:

A. MAP Representation

A map was obtained with the help of matplotlib which is a Python interface used to plot all nodes and ways with all description on the map. Nodes and ways are located on the plot with the help of latitude and longitude position of the particular node. The tool represents the map containing the brief information about the various places, ways enclosed in that region. From the above map, it can be concluded that the nodes in the region are properly mapped with mostly the required information but the ways still need some attention as many of ways are only marked with a 'highway' tag and there is no other information available for that like its name, maximum speed limit etc. Hence, the map containing the brief detail of this region still needs some attention from the point of completing the way information.

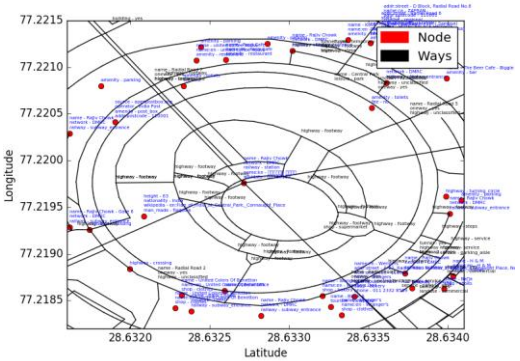

Figure 2. Map of selected area

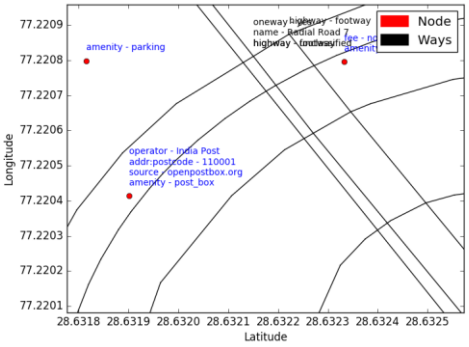

Figure 3. Zoomed Image of map

B. *Attribute Count*

This shows the count of the number of buildings, atms, ways and places in that region. If possible, it could then be compared to Google's map or some other map with the listed number of buildings, ATMs, ways, places and tolls of the same selected region to do some comments on the feature completeness. As Google map data is kept under professional supervision and is updated continuously, significant comments of that region completeness feature wise can be made by doing the comparison with the Google data. The figure 4 shows how many ways, building, ATM, and tolls exist in the selected area.

ATTRIBUTES

ATTRIBUTE	COUNT
Total Places	62
Total Ways	253
Total Building	9
Total ATM	0
Total Tolls	0

Figure 4. Attribute Count of the selected area

C. *Attribute completeness*

The attributes listed in the figure 5 are counted and stored and represented on the web interface in the form of a table. By this user can check the various areas of incompleteness in OSM data.

ATTRIBUTE COMPLETENESS

	ATTRIBUTE	COUNT
1. General Information on Study Area	Missing tags	649
	Ways without width	61
2. Points-Of-Interest	Shops without opening hours	6
	Worship places without religion	0
	Drive Through tag in pharmacy, bank, atm, or fastfood, postbox	0
	No of places with amenity = 'fast_food' or 'restaurant' or 'bar' or 'café' without cuisine tags	6
3. Geocoding	No. of POIs without names and housenumbers	13
	No. of building with missing tags Height and Levels	0
4. Routing and Navigation	No of Highways with missing tag 'maxspeed'	61
	No of Tolls with missing tag 'charge'	0
	No of Railway without crossing tags	2
	Ways with layer tag without tag bridge or tunnel	0

Figure 5. Attribute Completeness

D. *Number of nodes in each version*

A graph is plotted using the number of nodes against version wise. A new version of node is created only when some editing is done by any user to the node. More is number of nodes in the highest version, we could expect, more is the editing done and hence better is the result. The figure 6 shows the graph of the number of nodes in each version for the selected area.

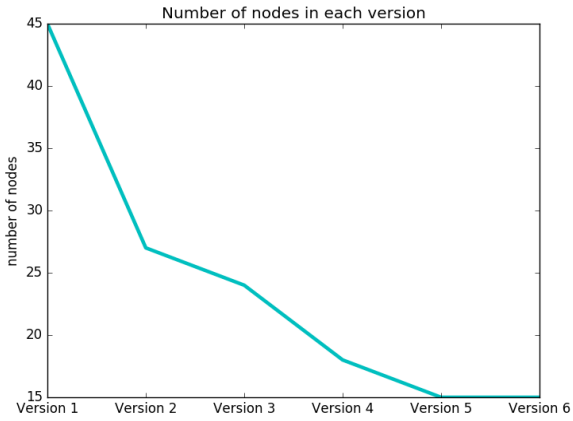

Figure 6. Number of nodes in each version

From the graph, it can be concluded that almost 66% of nodes has been edited atleast once and 33% of nodes has been edited six times to add more and more information positively. From this perspective, it can be concluded that the attribute completenesss is on increase and persistent.

E. *Node History*

The nodes are modified in the OSM with the time. If there are changes in some places, then a new version is created for a particular node. On the modification of the node, every time the current version of node is incremented by one. So when the new version is created for every node, then some

new information is added to the node. The node history details bar chart describes the mean number of tag detail of nodes version wise. Some version contains less number of tag details, and some contain a large number of details. So by this, we can check in which version the mean number of tags is more. If the mean number of tags is on increase version wise, the attribute completeness can be concluded as on increase. The bar chart in the figure 7 shows the mean number of tags in nodes versus version. It is showing the mean number of tags in nodes in each version.

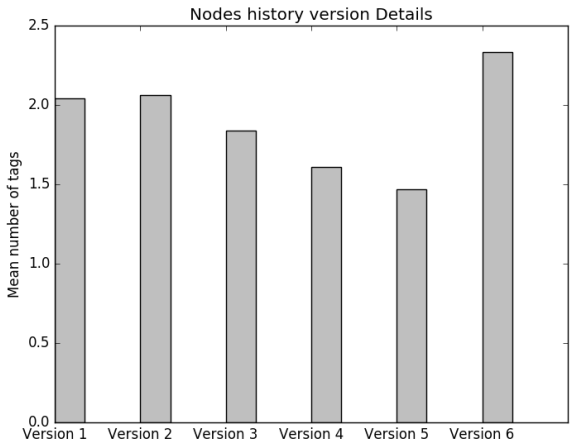

Figure 7. Node History Version details

The conclusion that we can withdraw fron the figure 7 is that the completeness of the region is increasing as the mean number of tags are increasing with each version. Hence, more and more information is being added for the edited nodes , making the OSM data more complete.

F. Count number of nodes, ways and buildings created in last five years

The evolution of a region over a specific time can help us in making the comments on completeness by providing a vision into the development and quality of that region. Hence, bar chart will allow us to interpret the qualitative and quantitative development of these regions. The nodes that has the 'building' tag will be added in the bar chart for building created in particular year. Every node and way contains time stamp which determines the creation of that node. Therefore, every node and way is checked for their time stamp and further, the time stamp is compared with the last five years. If the node belongs to the particular year, then the value of creation of the corresponding one is incremented.

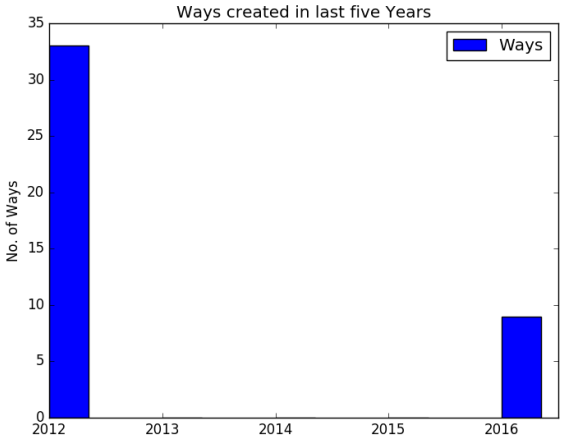

Figure 8. Ways created in last five years

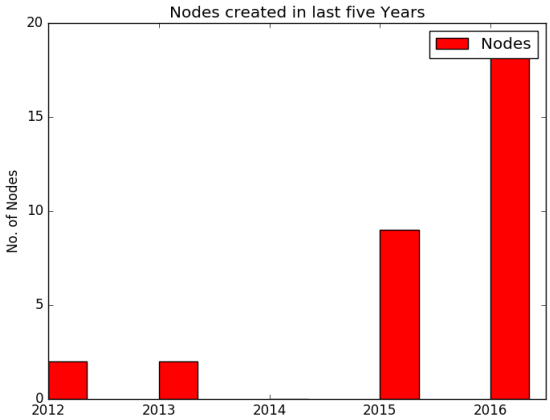

Figure 9. Nodes created in last five years

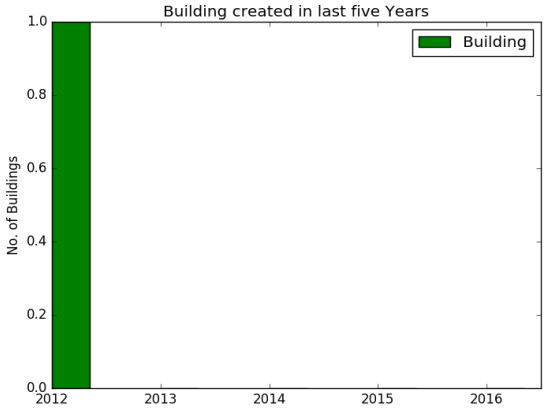

Figure 10. Buildings created in last five years

It is prevalent that in 2016, the number of nodes created is maximum as compared to the previous years. On the other

hand, maximum number of nodes and building were created in 2012 which might denotes a bulk import. Hence a significant decrease in the newly created building and node from 2012 could be commented as dataset for the two as near close to completion whereas since there is still the addition going on for others for the selected area.

G. Development

The development of the nodes and ways are examined by checking modification of node and ways in a particular year. Every year, most of the nodes and ways are developed and updated on the OSM. So every node and way contains time stamp which will determine the creation or modification of node or way in that year. The timestamp denotes the node, or a way is created/modified in that year which specifies the development. It means, if there are many nodes and ways which were modified in the recent year, the place is in developing mode. However, if some place has no creation/modification of node and way in the recent year will signify the incompleteness from the evolution point of view.

Development of Nodes in Last five years

Year	Development
Year 2012	1. Position : (28.6298659,77.2195453), Info : highway - traffic_signals, 2. Position : (28.6339853,77.2196152), Info : highway - turning_circle,
Year 2013	1. Position : (28.6323321,77.2207958), Info : amenity - toilets, fee - no, 2. Position : (28.6335193,77.2205682), Info : amenity - toilets, fee - no,
Year 2014	No Development in this year
Year 2015	1. Position : (28.6095117,77.2118249), Info : railway - station, network - DMRC, name - Udyog Bhavan, 2. Position : (28.6348111,77.2194381), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk, 3. Position : (28.6336132,77.2203351), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk, 4. Position : (28.6330177,77.2211762), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk, 5. Position : (28.6316175,77.2202911), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk, 6. Position : (28.631618,77.2192888), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk - Gate 6, 7. Position : (28.6337281,77.2187911), Info : railway - subway_entrance, network - DMRC, name - Rajiv Chowk, 8. Position : (28.6328022,77.2194015), Info : height - 63, wikipedia - en:Flag_of_India_at_Central_Park,_Connaught_Place, man_made - flagpole, nationality - India,
Year 2016	1. Position : (28.6314266,77.216835), Info : highway - traffic_signals, 2. Position : (28.6356983,77.2199282), Info : highway - traffic_signals, 3. Position : (28.6297527,77.2249956), Info : railway - station, network - DMRC, name - Barakhamba Road, 4. Position : (28.6312323,77.2224467), Info : highway - traffic_signals, 5. Position : (28.6325066,77.2177523), Info : cuisine - ice_cream, amenity - cafe, name - Cavanter's Milk Shake, 6. Position : (28.6314049,77.2204463), Info : note - nice view, amenity - restaurant, name - Connection Beer company, 7. Position : (28.6325905,77.218697), Info : name - postal office, 8. Position : (28.6336189,77.2211054), Info : note - good panir, amenity - fast_food, name - The Embassy, 9. Position : (28.6339741,77.2186255), Info : shop - shoes, name - Bata, 10. Position : (28.6319907,77.2204138), Info : operator - India Post, source - openpostbox.org, amenity - post_box, addr:postcode - 110001, 11. Position : (28.6332598,77.2194352), Info : tourism - viewpoint, name - Block A, 12. Position : (28.6324161,77.2210737), Info : name:en - united coffee house, amenity - restaurant, name - united coffee house, 13. Position : (28.6323227,77.2185561), Info : name:en - United Colors Of Benetton, shop - clothes, name - United Colors Of Benetton, 14. Position : (28.6322828,77.2184226), Info : name:en - United Colors Of Benetton, shop - clothes, name - United Colors Of Benetton, 15. Position : (28.6340328,77.2187608), Info : name:en - H & M, shop - clothes, name - H & M, 16. Position : (28.6339927,77.2190844), Info : amenity - bar, name - The Beer Cafe - Blazie,

Figure 11. Node development in last five years

Development of Ways in Last five years

Year	Development
Year 2012	1. Info : landuse - retail, name - Pallika Bazaar, 2. Info : highway - footway, 3. Info : highway - footway, 4. Info : highway - footway, 5. Info : highway - footway, 6. Info : highway - footway, 7. Info : highway - steps, 8. Info : highway - footway, 9. Info : highway - footway, 10. Info : highway - footway, 11. Info : highway - footway, 12. Info : highway - footway, 13. Info : highway - footway, 14. Info : highway - footway, 15. Info : highway - footway, 16. Info : highway - footway, 17. Info : highway - footway, 18. Info : layer - -1, tunnel - yes, highway - footway, 19. Info : highway - footway, 20. Info : highway - footway, 21. Info : highway - steps, 22. Info : leisure - park, name - Central Park, 23. Info : highway - footway, 24. Info : highway - footway, 25. Info : building - yes, 26. Info : highway - steps, 27. Info : highway - footway, 28. Info : highway - steps, 29. Info : layer - -1, tunnel - yes, highway - footway, 30. Info : highway - footway, 31. Info : highway - footway, 32. Info : layer - -1, tunnel - yes, highway - footway, 33. Info : service - parking_aisle, highway - service, 34. Info : highway - steps,
Year 2013	No Development in this year
Year 2014	No Development in this year
Year 2015	No Development in this year
Year 2016	1. Info : oneway - yes, name - Radial Road 1, highway - unclassified, 2. Info : oneway - yes, name - Radial Road 7, highway - unclassified, 3. Info : oneway - yes, name - Radial Road 6, highway - unclassified, 4. Info : oneway - yes, name - Radial Road 4, highway - unclassified, 5. Info : oneway - yes, name - Radial Road 3, highway - unclassified, 6. Info : railway - subway, tunnel - yes, voltage - 25000, electrified - contact_line, colour - 1e90ff, layer - -1, gauge - 1676, name - Delhi Metro Blue Line, frequ 7. Info : highway - footway,

Figure 12. Development of ways in last five years

The figure 11 is showing that there was development in the different years and the details of that nodes' development are listed in development column. The figure 12 is also showing that there was the major development of way in the year 2012 with the details given in development section. There was development going on in the year 2016. Hence, the region could be stated as still in the development phase.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

OpenStreetMap was found to encourage the development, growth, and distribution of free geographical data. The information from OSM is freely accessible for visualization, inquiry, download, and editing by open licenses, hence the OSM quality assessment is necessary. In this investigation, we analysed the limitations of the existing tools to measure the completeness of the OSM data. The research presented six different quality indicators to evaluate the OSM data quality using the OSM Api. Using the proposed parameters, a tool has been designed for measuring the completeness of the OSM data. The results of the study showed that the proposed tool is effective in providing complete information about the different attributes such as number of ways, buildings, atms and tolls in the selected area and their details in brief. It also provides information regarding the attribute completeness of the region. The above study in which a small region of New Delhi(India) has been considered is found as incomplete attribute wise in OSM on the basis of the results obtained. Many possible attributes related to features are missing and has a scope of updation. The Bar charts provided visualization of quantitative development of the region selected. On the basis of bar charts, the investigation revealed that the development is still going on in the region selected for the ways and nodes but is almost

completed for buildings. Hence, this enables arbitrarily OSM completeness assessment for any part of the world without using any ground truth reference dataset. But this study could only provide the approximate statements. For absolute statements, the reference dataset will be required.

B. Future Scope

Completeness of data is highly essential in OpenStreetMap. Therefore, in the future, the various factors could be explored that can further enhance the proposed tools. The number of features can be extended to ensure high accuracy and effectiveness of the tool in measuring the completeness of the OSM data. Also, as OSM quality heavily depends on the contribution made by user, their experience level, their activity and pace of contribution could also be considered. The modularity of the framework makes the extension easy for further addition of features and parameters.

REFERENCES

[1] J. N. D. Costa, "*Novel tool for examination of data completeness based on a comparative study of vgi data and official building datasets*", GEODETSKI VESTN, vol. 60, no. 3, pp. 495–508, 2016.

[2] A. Ballatore, M. Bertolotto, and D. C.Wilson, *"Geographic knowledge extraction and semantic similarity in openstreetmap*", Knowledge and Information Systems, vol. 37, no. 1, p. 61, 2013.

[3] T. Hayakawa, Y. Imi, and T. Ito, "Analysis of quality of data in openstreetmap", in Seventh IEEE International Conference on E-Commerce Technology (CEC'05) (2012), Hangzhou, China China, pp. 131–134, 2012.

[4] M. Haklay, "*How good is volunteered geographical information A comparative study of openstreetmap and ordnance survey datasets*", Environment and Planning B: Planning and Design, vol. 37, no. 4, pp. 682–703, 2010.

[5] C. Barron, P. Neis, and A. Zipf, "*A comprehensive framework for intrinsic openstreetmap quality analysis*", Transactions in GIS, vol. 18, no. 6, pp. 877–895, 2014.

[6] P. Mooney and P. Corcoran, "*Analysis of interaction and co-editing patterns amongst openstreetmap contributors*", Transactions in GIS, vol. 18, no. 5, pp. 633–659, 2014.

[7] C. Westrope, R. Banick, and M. Levine, "*Groundtruthing openstreetmap building damage assessment*", Procedia Engineering, vol. 78, pp. 29 – 39, 2014.

[8] H. Fan, A. Zipf, Qing Fu and Pascal Neis, "*Quality assessment for building footprints data on OpenStreetMap*", International Journal of Geographical Information Science, vol. 28, Issue 4, pp. 700-719, 2014.

[9] S. S. Sehra, J. Singh, and H. S. Rai, "A systematic study of openstreetmap data quality assessment", in Information Technology: New Generations (ITNG), 2014 11th International Conference, Las Vegas, NV, USA, pp. 377–381, 2014.

[10] V. Antoniou and A. Skopeliti, "*Measures and indicators of vgi quality: An overview*", ISPRS Internationa Society for Photogrammetry and Remote Sensing, vol. II-3/W5, pp.345–351, 2015.

[11] P. Neis, M. Goetz, and A. Zipf, "*Towards automatic vandalism detection in openstreetmap*", ISPRS International Journal of Geo-Information, vol. 1, no. 3, p.315, 2012.

[12] M. S. Fox and C. J. Pettit, "On the completeness of open city data for measuring city indicators", in Smart Cities Conference (ISC2). Guadalajara, Mexico: IEEE, pp. 1-6 , 2015.

[13] J. Vuurstaek, G. Cich, L. Knapen, A.-U.-H. Yasar, T. Bellemans, and D. Janssens, "*Gtfsbus stop mapping to the osm network*", Procedia Computer Science, vol. 109, pp. 50–58, 2017.

[14] J.F. Girres and G. Touya, "*Quality assessment of the french openstreetmap dataset*", Transactions in GIS, vol. 14, Issue 4, pp. 435–459, 2010.

[15] B.A. Johnson and K. Iizuka, "*Integrating openstreetmap crowdsourced data and landsat time- series imagery for rapid land use/land cover (lulc) mapping: Case study of the laguna de bay area of the philippines*", Applied Geography, vol. 67, pp. 140–149, 2016.

[16] C. Heipke, "*Crowdsourcing geospatial data*", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, pp. 550–557, 2010.

[17] M. Over, A. Schilling, S. Neubauer, and A. Zipf, "*Generating web-based 3d city models from openstreetmap: The current situation in germany*", Computers, Environment and Urban Systems, vol. 34, no. 6, pp. 496 – 507, 2010.

[18] M. Ruta, F. Scioscia, D. D. Filippis, S. Ieva, M. Binetti, and E. D. Sciascio, "*A semanticenhanced augmented reality tool for openstreetmap poi discovery*", Transportation Research Procedia, vol. 3, pp. 479–488, 2014.

[19] F. Scioscia, M. Binetti, M. Ruta, S. Ieva, and E. D. Sciascio, "*A framework and a tool for semantic annotation of fPOIsg in openstreetmap*", Procedia - Social and Behavioral Sciences, vol. 111, pp. 1092 – 1101, 2014.

[20] C. Bittner, "*Openstreetmap in israel and palestine game changer or reproducer of contested cartographies?*", Political Geography, vol. 57, pp. 3448, 2017.

[21] J. J. Arsanjani, "*An assessment of a collaborative mapping approach for exploring land use patterns for several european metropolises*", International Journal of Applied Earth Observation and Geoinformation, vol. 35, pp. 329–337, 2015.

[22] J. Schellekens, R. Brolsma, R. Dahm, G. Donchyts, and H. Winsemius, "*Rapid setup of hydrological and hydraulic models using openstreetmap and the srtm derived digital elevation model*", Environmental Modelling & Software, vol. 61, pp. 98–105, 2014.

[23] P. Zhao, T. Jia, K. Qin, J. Shan, and C. Jiao, "*Statistical analysis on the evolution of openstreetmap road networks in beijing*", Physica A: Statistical Mechanics and its Applications, vol. 420, pp. 59 – 72, 2015.

[24] M F Goodchild 2007 "*Citizens as sensors: The world of volunteered geography*", GeoJournal , vol. 69, pp. 211–21, 2007.

[25] G. Sivaiah, PK. Rao, "*Providing Efficient Driving Directions Using GPS and Driver`s Ability*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.9, pp.18-21, 2014.

Authors Profile

Mrs. Sonali Arora pursed her Bachelor of Science from RIMT University, Mandi Gobindgarh in 2012 and Master of Science from GNDEC in year 2017. She has four years of IT experience and has worked for Steria, Noida as Junior Software Engineer for one year and for Capgemini, Pune as Software Engineer for three years. She has worked for credit card application of Barclays Bank, United Kingdom. Her main research work focuses on OpenStreetMap, Barclays Bank Credit Card Application development, Programming in Java, Big Data Analytics, Handling Self Serve Account, Data Mining.

Mr Sukhjit Singh Sehra pursed Bachelor of Science in 2002 and Master of Science in year 2006. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, GNDEC, Ludhiana since 2006. His main research work focuses on Spatial Databases, Data Mining, Grid/ Cloud Computing, OpenStreetMap. He has 10 years of teaching experience.

Mrs Sumeet Kaur Sehra pursed Bachelor of Science in 2002 and Master of Science in year 2006. She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, GNDEC, Ludhiana since 2006. His main research work focuses on Spatial Databases, Data Mining, Grid/ Cloud Computing, OpenStreetMap, Software Engineering, Machine Learning Algorithms. She has 10 years of teaching experience.

