Assessment of OpenStreetMap Data - A Review

Sukhjit Singh Sehra
Assistant Professor
GNDEC, Ludhiana, India

Jaiteg Singh
Associate Professor
CIET, Punjab, India

Hardeep Singh Rai
Professor
GNDEC, Ludhiana, India

ABSTRACT

The meaning and purposes of web has been changing and evolving day by day. Web 2.0 encouraged more contribution by the end users. This movement provided revolutionary methods of sharing and computing data by crowdsourcing such as OpenStreetmap, also called "the wikification of maps" by some researchers. When crowdsourcing collects huge data with help of general public with varying level of mapping experience, the focus of researcher should be on analysing the data rather than collecting it. Researchers have assessed the quality of OpenStreetMap data by comparing it with proprietary data or data of governmental map agencies. This study reviews the research work for assessment of OpenStreetMap Data and also discusses about the future directions.

General Terms:

Assessment, OpenStreetMap

Keywords:

Openstreetmap, Crowdsourcing, Voltereed Geographic Information

1. INTRODUCTION

The Web started with web 1.0 or "read only" web, but Tim O'Reilly [1] discussed the concept of Web 2.0 or "read-write" web. The client side technologies such as Ajax and JavaScript framework are used for Web 2.0 development [1, 2]. Web 2.0 encourages greater collaboration among internet users and other users, content providers, and enterprises [3]. This movement provided revolutionary new methods of sharing and computing data by crowdsourcing movement similar to wikipedia [4, 5, 6]. In regard to the geographical data the crowd-sourced movement is known as VGI (volunteered geographic information) [8], others call it collaborative mapping [9], so it is a special case of this web phenomenon and has been applied in many popular websites such as Wikimapia, OpenStreetMap(OSM), GoogleMap, Flickr [10]. The two major data providers are NavTeq and TeleAtlas. However, these data are costly, quickly outdated and restricted to specific areas covered by the data acquiring companies. Large companies have invested large sums of money to purchase smaller companies to acquire their data e.g. in 2007, Nokia acquired NavTeq, in 2006, Microsoft acquired the Imagery and Remote Sensing Company Vexcel [12].

The website of Openstreetmap is a collaborative project to create a free editable map of the world, using crowdsourced approach. Two major driving forces behind the establishment and growth of OpenStreetMap have been restrictions on use or availability of map information across much of the world and the advent of inexpensive portable satellite navigation devices. GPS enabled smartphones based on iOS (72 millon Iphones worldwide) or Android (One billion users worldwide) [15] made easy for users to contribute to crowd sourced openstreetmap. OpenStreetMap is based on the concept of crowdsourcing, also called wikification of GIS [13], which encourages the volunteers worldwide to contribute through the collection of geographic data. The data of OpenStreetMap is useful because firstly, the data is completely free with an open content licence. Secondly, it is constantly being updated by the subscribed users who can also add points of interest important to them. OpenStreetMap has been used during earthquake situation in Haiti named as Haiti Crisis Map. Finally, OpenStreetMap has the potential to establish volunteers from all over world including less developed regions, where obtaining data can be difficult for most commercial mapping companies [14]. Till July, 2013 nearly 1,300,000 have registered as users/contributors to OpenStreetMap [11], fig 1 show the graph. To compete with OpenStreetMap, Google introduced a tool called Map Maker in 2008, that enabled users to contribute data themselves. This tool was only available for areas with no or little commercial data coverage, e.g. India, Pakistan, Iceland and within a short time, large areas were mapped in this crowdsourcing manner [12].

Next section discusses about the assessment work on Openstreetmap. Nearly every researchers have compared the OpenStreetMap Data with actual ground data and some compared it with the proprietary data sets. Last section is conclusion and future directions.

2. QUALITY ASSESSMENT PARAMETERS

For OpenStreetMap to replace proprietary maps, the quality of map data has to assessed, so many researchers started comparing OSM data with ground data available with different governmental agencies. The Geographical information from different sources may be compared based following parameters [10, 14, 18]:-

(1) ineage: - It describes the record of the spatial data, which includes the detailed description of the source material from which the data were derived, the methods of derivation, the dates of the source material, and all transformations involved in producing the final digital files.

[(2)] ositional Accuracy: - Positional accuracy defines the closeness of locational information to the true position i.e. the absolute and relative accuracy of the positions of geographic features. The Absolute accuracy is the closeness of the co-ordinate values in a dataset to values accepted as or being true and the Relative accuracy is the closeness of the relative positions of features to their respective relative positions accepted as or being true.

[(3)] ompleteness: - It describes the degree to which geographic features, their attributes and their relationships are included or not included in spatial dataset. In addition it also include information on selection criteria, definitions used, and other relevant mapping rules.
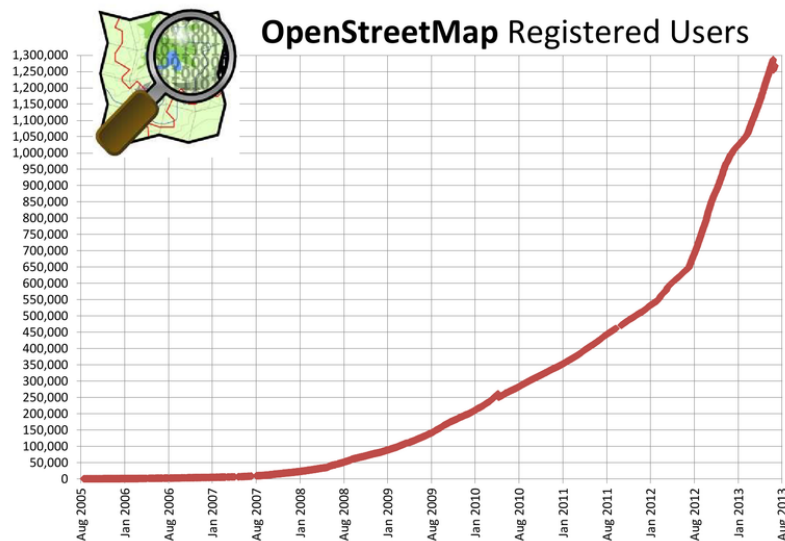
Fig. 1. Registered Users

Some researcher also analysed the spatial data quality using parameters in addition to discussed above, which are Thematic Accuracy, Temporal Accuracy, Logical consistency, Semantic accuracy, Usage/purpose/constraints, Variation in quality, Metaquality, and Resolution. Often one dataset may be superior to other datasets in one, but not all aspects.

3. ASSESSMENT WORK

OpenStreetMap is based on the concept of crowdsourcing and founded by Steve Coast from UK in 2004. OpenStreetMap produces huge spatial data, with less effort. and researchers are working on the devising method to use the data rather than collecting the data. OpenStreetMap produces labelled data, When labeled data is easy to come by, the focus of the researcher would be on working with the labelled data rather than collecting it. This impacts not only which problems researchers choose to work on, but also the learning methodology they use to approach them.

In this sense, crowdsourcing provides a two-part cost savings: greater ability to realize traditionally-claimed savings of active learning, as well as reduced cost of crowd annotation vs. traditional annotators. A second important benefit will again be the implications for use of labeled vs. unlabeled data for training when labeled data is plentiful. Instead of comparing to past supervised learning curves, researchers may instead consider past learning curves for active learning, which will be steeper in comparison. In many prior studies, it has been concluded that there is greater potential for active learning than supervised learning to benefit from crowdsourcing [16, 17].

Haklay, in 2008, [18] analysed OSM data compared Great Britain, and ordnance survey (OS) geodata with OSM data. In 2009, Ather [14] extended this work to the OS Master Map for selected parts of London. He additionally compared completeness of road names. Then, in 2010, Haklay [19] in his research work buffered British Ordnance Survey data to determine what percentage of the OSM roads were covered. A commonly applied technique for matching different road networks is graph matching [20]. The analysis of Germany started with comparison of commercial mulinet proprietary map data from TomTom [21, 22] compared with street map data from different proprietary geodata providers.

In 2011, Ciepluch [30] discussed that not everyone contributes data of the same quality, due to lack of practice and knowledge which can be improved by practice and experience in map making. Ludwig [22] described a methodology to compare OSM street data with Navteq for all populated roads in Germany. The methodology was based on a matching between the street objects of OSM and Navteq adopted from [23] which allows for object-wise comparison of geometries and thematic attributes. Finally, they calculated relative quality measures: relative object completeness, relative attribute completeness, difference in speed limits and positional differences. Another researcher [24] statistically analysed the routing process using OpenStreetMap road data of the inner city of Hamburg. A similar approach was used in France to analyse OSM data [25]. The results of this research showed the advantage and flexibility, but also concluded the problem of the heterogeneity of the data specifically for France. In same year, the first study that analysed the quality of OSM outside of Europe were conducted [26]. In this research work the OSM project data had been compared with proprietary data from TomTom (TeleAtlas) and Navteq for the entire state of Florida (USA) and four specific cities within the USA. In comparison to the results for Germany or England, the discrepancies between the rural and urban areas in the USA showed an opposite tendency. In Florida, the rural data was, in parts, even more complete than that of the proprietary datasets in the relative comparison conducted. Zielstra [27] compared the amount of pedestrian-related data between freely available sources, i.e., OSM and/or TIGER, and proprietary providers, i.e., Tele Atlas Multinet and/or Navteq Discover Cities, and analysed its effect on modeling ?spatial aspects of transit accessibility" for pedestrian in five US and four German cities. They concluded that integration of pedestrian-only segments can lead to a more realistic assessment of service areas when compared to using networks that contain only streets that are passable by cars and that the assessment of VGI data quality, especially OSM data, is an ongoing issue of high importance for successful geo-applications [7, 19].

Other analyst in 2012 [28] assessed the effect of network data integration from different sources on the length of computed shortest paths for pedestrians. Their results showed that the combined use of network datasets significantly reduces shortest path distances compared to the use of single datasets. They concluded that data integration leads to an increased value for users of

pedestrian routing applications but that combining OSM and other commercial datasets cannot be considered for implementation due to current licensing issues. Neis [29] assessed the completeness of the OSM street network via a relative comparison (street network length, no. streets without names, no. turn restrictions) between OSM and a commercial dataset provider (Tom-Tom formerly known as Tele Atlas). They noted though that for comparison the TomTom dataset is suitable only for street network data for car-specific navigation. They also evaluated logical consistency using an internal test, whereby topological and thematic consistency is determined. Concerning turn restrictions (176,000 in TomTom; 21,000 in OSM in Germany in June 2011) they discussed that although the number of turn restrictions available in the OSM dataset is continually increasing, it will probably take several more years before OSM achieves the same level as TomTom, based on the current status and development?. Apart from England, no studies have been conducted to date over a period of several years and for an entire country [19].

In 2013 many researchers have been aggressively working in the area of assessment of OpenStreetMap, but research work for India has not initiated yet.

4. CONCLUSIONS & FUTURE WORK

This review paper concludes that OpenstreetMap is generating huge dataset with the help of non-commercialised users, with varying level of experience. So the assessment becomes vital to give maturity to OpenStreetMap data. The results from the finding show that because of varying level of user experience the data is not error free, and also mapped areas depend upon the contribution by the users. But general trend is that number of absolute and relative errors is falling.

The discussed approaches in this paper are offline approaches of checking the data and correcting the OSM data afterwards. Another method which still needs attention from the researchers community is online quality check or anomaly detection engine, that can check for the quality of the map while it is being uploaded by the user. The idea may be to create automated model that can spot mistakes. In order to do this some machine learning techniques with several user parameters being: the age of the user, the number of edits, what happens in a changeset etc can be taken into consideration as discussed by Neis [29]. Further some of the researchers have compared data proprietary data sets and given their results. The ground reality of those proprietary data sets may also be checked.

5. REFERENCES

[1] O' Reilly, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, OReilly Media: Cambridge, MA, USA, 2008. `http://oreilly.com/web2/archive/what-is-web-20.html` (Acessed on 13/3/13).

[2] Anderson, P. What Is Web 2.0? Ideas, Technologies and Implications for Education,Joint Information Systems Committee: Bristol, UK, 2007. `http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf` (Acessed on 13/3/13).

[3] Hudson-Smith, A.; Batty, M.; Milton, R.; Batty, M. NeoGeography and web 2.0: Concepts, tools and applications, in *Journal of Location Based Services*,Vol. 3, No. 2, 2009, 118–145.

[4] Goodchild, M. NeoGeography and the nature of geographic expertise. in *Journal of Location Based Services*,Vol. 3, No. 2,2009 82–96.

[5] Haklay, Muki, Alex Singleton, and Chris Parke Web mapping 2.0: The neogeography of the GeoWeb, in *Geography Compass 2.6*,2008, 2011-2039.

[6] Howe, Jeff The rise of crowdsourcing, in *Wired Magazine*, Issue 14.06,2006, 1–4.

[7] Goodchild, M.F. Spatial Accuracy 2.0. in *Proceeding of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Shanghai, China*, , 2008, 25–27.

[8] Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0 in *International Journal of Spatial Data Infrastructures Research*, 2007, Vol. 2, 24–32.

[9] Fischer, F. Collaborative mappingHow wikinomics is manifest in the geo-information economy in *GEOInformatics*, Volume 11, No 2, 2008, 28–31.

[10] Kounadi, O. Assessing the quality of OpenStreetMap data *MSc. Dissertation, University College of London Department of Civil, Environmental And Geomatic Engineering*, August 2009.

[11] The Data Stat of OpenStreetMap Users `http://wiki.openstreetmap.org/wiki/Stats` (Accessed 23/7/2013).

[12] Schmidt, M.; Weiser, P., Web Mapping Services: Development and Trends, in Springer Lecture Notes On Online Maps with APIs and WebServices, 2012, 13–21.

[13] Sui, D.Z. The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS, *Computers Environment and Urban Systems*, 2008, 32, 1–5.

[14] Ather, A. A Quality Analysis of OpenStreetMap Data M.E. Thesis, University College London, London, UK, May 2009.

[15] Android to command nearly half of worldwide smartphone operating system market by year-end 2012. URL `http://www.gartner.com/newsroom/id/1622614`, Accessed on 19/5/2013.

[16] Lease, M. On Quality Control and Machine Learning in Crowdsourcing in *Proceeding of Human Computation, AAAI Workshop*, August 2011, 97–102.

[17] Ambati, V., Active Learning and Crowdsourcing for Machine Translation in Low Resource Scenarios in *Thesis, Kevin Knight, University of Southern California, ISI*, 2012.

[18] Hakley, M. How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England in *Environment and Planning B: Planning and Design*, August 2008.

[19] Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets in *Environment and Planning B: Planning and Design*, 2010, 37(4), 682–703.

[20] Zhang, M. and Meng, I. An iterative road-matching approach for the integration of postal data in *Computer, Environment and Urban Systems*, 2007, 31, 597–615.

[21] Zielstra, D.; Zipf, A. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany in *Proceedings of 13th AGILE International Conference on Geographic Information Science, Guimar Portugal*, May 2010, 10–14.

[22] Ludwig, I.; Voss, A.; Krause-Traudes, M. A Comparison of the Street Networks of Navteq and OSM in Germany in *Advancing Geoinformation Scienced of a Changing World Lecture Notes in Geoinformation and Cartography 2011, published by Springer*, 65–84.

[23] Walter; Volker; Fritsch, D., Matching spatial data sets: a statistical approach in *International Journal of Geographical Information Science*, 13.5 (1999), 445–473.

[24] Fessele, M.: Poplin, A. Statistical analysis of routing processes using OpenStreetMap road data of Hamburg with different completeness of information about one-way streets in *Proceedings of GeoValue'2010*, 2010, 87–92.

[25] Girres, J.F.; Touya, G Quality assessment of the French OpenStreetMap dataset in *Transaction in GIS*, 2010, 14, 435–459.

[26] Zielstra, D.; Hochmair, H.H. Digital street data: Free versus proprietary in *GIM International*, 2011, 25, 29–33.

[27] Zielstra, D.; Hochmair, H.H. A comparative study of pedestrian accessibility to transit stations using free and proprietary network data in *Transportation Research Record: Journal of the Transportation Research Board*, 2011, 2217, 145–152.

[28] Zielstra, D.; Hochmair, H.H. Comparison of Shortest Path Lengths for Pedestrian Routing in Street Networks Using Free and Proprietary Data in *Proceedings of Transportation Research Board - 91st Annual Meeting, Washington, DC, USA*, 22–26, January 2012.

[29] Neis, P.; Zielstra, D.; Zipf, A., The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011 in *Future Internet*, 2012, 4(1), 1-21, doi:10.3390/fi4010001.

[30] Ciepluch, B.,;Mooney, P.,; Jacob, R.,; Winstanley, A., Sketches of Generic Framework for Quality Assessmenft of Volunteered Geographical Data in *IEEE Geoscience and Remote Sensing Society (GRSS)*, 2011 , 1–5.