Towards intrinsic Quality Analysis of OpenStreetMap Datasets

Christopher Barron
University of Heidelberg, GIScience Research Group
Berliner Straße 48,
69120 Heidelberg, Germany,
christopher.barron@geog.uni-heidelberg.de

Pascal Neis
University of Heidelberg, GIScience Research Group
Berliner Straße 48,
69120 Heidelberg, Germany,
neis@uni-heidelberg.de

Alexander Zipf
University of Heidelberg, GIScience Research Group
Berliner Straße 48,
69120 Heidelberg, Germany,
zipf@uni-heidelberg.de

1 Introduction

The increasing availability of voluntarily and collaboratively collected geodata, in particular *OpenStreetMap* (OSM), leads to numerous scientific studies with a focus on the evaluation of this data [1, 2, 3, 4]. A number of methods have been applied and enhanced to evaluate different data quality aspects. In most cases, data is evaluated and compared with administrative and commercial datasets. However, accessibility to the high quality and commercial datasets are limited due to the costs and licensing restrictions. Therefore, indicators or tools which do not require a reference dataset could be an alternate approach. For this purpose so-called "intrinsic indicators" focussing only on the data itself are suited. Yet, no framework in literature is evident, which combines different indicators and methods. This study attempts to bridge this gap.

2 The iOSMAnalyzer Tool

iOSMAnalyzer is a tool developed by the authors for intrinsic OSM quality analysis based on the data's full history. Several methods from OSM quality analysis in scientific literature are picked and developed further.

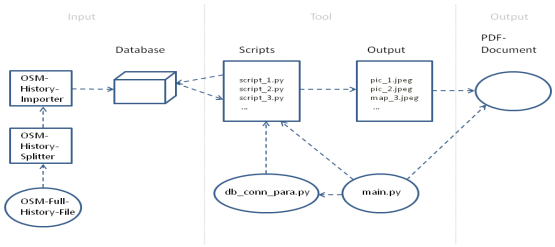
One example is the analysis of road intersections as described by [2]. Concerning their positional accuracy, the authors compare the distance of distinct road intersections of OSM data and commercial reference datasets to each other. For this study the method was enhanced. By analysing the road intersections for different timestamps of the OSM data, their distance and angle to each other can be visualized on a polar scatterplot. Subsequently, possible offsets between various timestamps of the data can be identified, caused by eg. displaced aerial imagery or major displacements of the road network.

A further example is the completeness of the OSM road network length [3]. In an intrinsic approach, the completeness approximately can be evaluated by analyzing the total monthly rate of growth within different road categories (eg. motorways, secondary/tertiary or residential roads). A minor monthly increase or even stagnation can indicate the appropriate road category of being "close to completion".

The focal point in this investigation is the integration of several intrinsic indicators for data quality in consideration of an OSM-Full-History-Dump, because prior studies mainly focus on temporal data snapshots.

In this holistic approach an OSM-Full-History-Dump for a freely selectable area is analyzed through considering 27 different quality indicators. The target area is clipped through the OSM-History-Splitter [5] and imported to a PostgreSQL/PostGIS database by the OSM-History-Importer [6]. Figure 1 illustrates the entire architecture and workflow of the developed tool.

Figure 1: Architecture of the iOSMAnalyzer.



In order to evaluate the data in terms of "Fitness for Purpose" the results are divided into the following use cases: "General Area Information", "Routing & Navigation", "Address-Search", "Points of Interest-Search", "Map-Applications" and "User -Information & -behavior". The subsequent quality indicators are being considered in *iOSMAnalyzer*:

- General Area Information:
  - Development of points, lines & polygons
  - Actuality of the dataset
  - Comparison of newly created and edited objects
  - Quantitative analysis in respect to the attributes of OSM features
  - Syntactic attribute accuracy
  - Positional accuracy of road intersections from different timestamps
- Routing & Navigation
  - Development of the total road network length
  - Development of the syntactic attribute accuracy of roads
  - Actuality of the road network
  - Logical consistency of the total road network
  - Roads without a name or route number
  - Grid-based analysis of the positional accuracy of the total road network

- Address-Search
    - Increase/decrease of address information over time
    - Completeness of house numbers tagged to buildings
- „Points of Interest“- Search
    - Development of POIs & average number of attributes
    - Attributive completeness of several POI categories
- Map-Applications
    - Development of the geometry representation of polygons
    - Untouched points, lines & polygons
    - Development of invalid polygons
    - Logical consistency of landuse polygons
    - Quantitative development of selected polygons
- User -Information & -behavior
    - Number & activity of contributors
    - Distribution of contributions
    - User profiles

For each use case, several preliminary developed Python-scripts automatically generate tables, maps and diagrams. The results are plotted into a single pdf-file which gives an overall view of several quality indicators within an area of choice.

3 Summary

A number of indicators and methods where presented which enable the analysis of several quality aspects of an OSM-Full-History-Dump. Relating to various other studies on OSM data quality, the benefit of the *iOSMAnalyzer* can be seen in the intrinsic and holistic approach without any ground-truth reference data. Furthermore, besides the intrinsic approach, all deployed software components are open source so any area of interest can be applied. Finally, the modular structure of the tool allows a simple integration of new additional components.

An important issue for further work is the consideration of relations which are not taken into account yet because they are not supported by the OSM-History-Importer.

References

[1] M. Haklay. How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. In: *Environment and Planning B: Planning and Design*, 37 (4), pages 682-703, 2008.

[2] M. Helbich, C. Amelunxen, P. Neis and A. Zipf. Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. In: J. Strobl et al., *Angewandte Geoinformatik 2012*, pages 24-33. VDE Verlag GmbH, Berlin/Offenbach, 2012.

[3] P. Neis, D. Zielstra, A. Zipf. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007-2011. In: *Future Internet,* 2012 (4), pages 1-21, 2012.

[4] D. Zielstra and A. Zipf. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. AGILE 2010. The 13th AGILE International Conference on Geographic Information Science. Guimarães, Portugal, 2010.

[5] https://github.com/MaZderMind/osm-history-splitter (24 March 2013).

[6] https://github.com/MaZderMind/osm-history-renderer (24 March 2013).