Use of Spatial and Temporal Editing Patterns for the Evaluation of OpenStreetMap data

Gröchenig, S.(1), Paulus, G.(1), Rehrl, K. (2) & Hochmair H. (3)

(1) Department of Geoinformation & Environemtal Technologies, Carinthia University of Applied Sciences, Europastrasse 4, A-9524 Villach, Austria
(2) Salzburg Research Forschungsgesellschaft m.b.H., Jakob Haringer Straße 5/3, 5020 Salzburg, Austria
(3) Fort Lauderdale Research and Education Center, University of Florida, 3205 College Ave., Davie, FL 33314-7799, USA

Motivation

Volunteered Geographic Information (VGI) (Goodchild, 2007) has become a widely used alternative to commercial datasets for a variety of geo-applications. The OpenStreetMap (OSM) project, which has the goal to create a detailed map of the world based on VGI in vector data format, is one of the most prominent Web 2.0 applications. The OSM database is populated through voluntary data contributors and therefore not governed by an authoritative agency. Since many of the voluntary contributors do not have a formal training in geo-data handling, data validation is crucial to warrant usability of OSM data for geo-applications. Previous research assessed OSM data quality primarily through comparison with commercial or governmental reference data sets (Hacklay 2010; Zielstra and Hochmair 2011). This work presents a novel intrinsic approach for quality assessment, which means that only OSM data itself, more specifically, the editing history of features, is used for the evaluation.

Methodology

The algorithm is based on the activity theory defined by Kuuttii (1996), who introduced the operation – action – activity hierarchy. After downloading the OSM features and their history files for a selected area the basic operations performed on features are identified. This includes (1) node coordinate changes, modifications of (2) the node list for ways and (3) the member list of relations, and (4) updated tags. Also special operations, such as “way split” or the “feature re-creation” are supported. Next, the basic operations are sorted by time and mapper and aggregated to actions based on a set of action definitions. For example, the “Create Line” action must have (1) a “create way”, (2) at least two “add node”, (3) and some “add attribute” operations. Each action is linked to a quality parameter (positional accuracy, thematical accuracy, completeness, logical consistency and temporal accuracy) and has a quality value for that parameter. This link represents the activity level in the mentioned activity theory. An action with fewer aggregated operations has a lower quality value than an action with more operations. Eventually, the history for every map feature is extracted, based on which the quality measure is determined. The algorithm sums up all quality values for each feature in order to get an idea of the data quality of the selected area.

Results

The algorithm produces an output which shows how many actions are found for which feature type (e.g., highway, amenity, landuse, or building). Since each action is linked to a quality parameter, the quality of different feature types, measured along the five quality parameters, can be assessed for the selected area. The algorithm allows setting a time range for analyzed edits, which can be used to identify temporal patterns in quality improvement for different feature types.

References

Goodchild M., 2007, Citizens as sensors: The world of volunteered geography, *GeoJournal*, 69 (4), 211-221

Haklay, M. (2010). How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and Planning B, Planning and Design*, 37 (4), 682 – 703.

Kuuttii, K., 1996. Activity theory as a potential framework for human computer interaction research. *In Context and consciousness: Activity theory and human-computer interaction*, ed. B.A. Nardi, 17-44. Cambridge, MA: The MIT Press.

Zielstra, D. and Hochmair H.H., 2011, A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data, Jo*urnal of the Transportation Research Board*, 2217, 145–152