

IMPROVING VOLUNTEERED GEOGRAPHIC DATA QUALITY USING SEMANTIC SIMILARITY MEASUREMENTS

A. Vandecasteele and R. Devillers

Department of Geography, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3X9
(a.vandecasteele, rdeville)@mun.ca

KEY WORDS: Volunteered geographic information, Data quality, Semantic similarity, Semantic heterogeneity, OpenStreetMap

ABSTRACT:

Studies have analysed the quality of volunteered geographic information (VGI) datasets, assessing the positional accuracy of features and the completeness of specific attributes. While it has been shown that VGI can, in some context, reach a high positional accuracy, these works have also highlighted a large spatial heterogeneity in positional accuracy, completeness but also with regards to the semantics of the objects. Such high semantic heterogeneity of VGI datasets becomes a significant obstacle to a number of possible uses that could be made of the data.

This paper proposes an approach for both improving the semantic quality and reducing the semantic heterogeneity of VGI datasets. The improvement of the semantic quality is achieved by automatically suggesting attributes to contributors during the editing process. The reduction of semantic heterogeneity is achieved by automatically notifying contributors when two attributes are too similar or too dissimilar. The approach was implemented into a plugin for OpenStreetMap and different examples illustrate how this plugin can be used to improve the quality of VGI data.

1. INTRODUCTION

With the increasing availability of geolocation devices and the development of more user-friendly software, expertise in cartography is no longer necessary to collect, publish and share geographic information (GI) (Goodchild and Li, 2012; Haklay, 2013). These developments have led to the emergence of a new phenomenon named Volunteered Geographic Information (VGI) (Goodchild, 2007). VGI, a geographic type of user generated content, has been described by Goodchild (2007) as resulting from "the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information." VGI opened the door to an alternative to traditional data producers by providing citizens with tools and web services to describe and share their geographic knowledge (Flanagin and Metzger, 2008).

Despite the great potential of VGI, it has not been widely adopted by the geographic information system (GIS) community mainly due to the absence of detailed information about its quality (Flanagin and Metzger, 2008; Elwood et al., 2012). If the quality evaluations made by several studies (Girres and Touya, 2010; Haklay et al., 2010; Ludwig et al., 2011) have documented relatively high overall positional accuracy of VGI data, they have also identified challenges. In terms of coverage, some areas can be well mapped while others not (Goodchild, 2007). In terms of attributes, VGI data can be unclear and their quality can vary spatially and through time (Ballatore et al., 2012). Different names (e.g., forest and wood) can sometime represent similar geographic phenomena (synonymy) while the same name can sometimes be used to describe different geographic phenomena (polysemy) (Ballatore et al., 2012; Mooney and Corcoran, 2012a).

The nature of VGI compared to typical cartographic products makes the use of traditional data quality evaluation methods difficult and new data quality metrics have been developed (van Exel and Dias, 2011; Keßler and Groot, 2013). While positional accuracy assessments of VGI have received significant attention, fewer efforts have looked at the quality of

VGI semantics (Mooney and Corcoran, 2012a). Moreover, no studies have attempted to enhance the quality of VGI during the editing process. This paper proposes using semantic similarity measurements to enhance users' experience by automatically suggesting attributes or by detecting attributes that are not related to each other. This approach has been implemented and tested by developing a semantic editor plugin for the OpenStreetMap (OSM) VGI project. Started in 2004, OpenStreetMap* is the most popular example of a VGI project. OSM aims to create a free geospatial database crowdsourced by volunteers all over the world. Since the beginning of the project, OSM has grown rapidly and has seen its community reach 1 million volunteers in early 2013. At the time of writing this paper the OSM project has more than 3 billion GPS points, 1.8 billion nodes and 175 million ways (i.e., polylines and polygons)**.

This paper is structured as follows. Section 2 presents relevant work about quality evaluation of VGI dataset. Section 3 describes the principles of semantic similarity measurements. Section 4 presents the prototype and some examples of its use.

2. QUALITY EVALUATION OF VGI DATASET

VGI allows people with little knowledge in GI to contribute to the creation of maps that are made publicly available. In a context where mapping was traditionally done by professional cartographers, such a change has been met with significant initial scepticism on the quality of VGI data which has led to a number of academic studies. However, the nature of VGI makes classical data quality elements (e.g., completeness and positional accuracy) and their measurements, as defined by the ISO/TC 211 (ISO/TC 211, 2002), of limited use in some cases. Data quality can also be described in terms of fitness for use, also named external quality (Devillers et al., 2010). New metrics for assessing external quality of VGI may have to be proposed to capture specific properties of VGI data. Some of the possible missing quality elements include measurements of

* <http://www.openstreetmap.org/>

** http://www.openstreetmap.org/stats/data_stats.html

the spatial and semantic heterogeneity of data, a problem that did not exist with most traditional maps that had precise specifications allowing the production of relatively homogeneous data products.

2.1 Measuring VGI data quality

Despite the relative novelty of VGI, several research studies have analysed the quality of OSM dataset. One of the popular and classical approaches used to assess the overall quality of VGI dataset is to compare VGI to an authoritative dataset of the same area that acts as ground-truthing data. These comparisons used different measurements, such as the length of the features or the difference between the attribute values. These analyses have been done in different countries such as England (Haklay et al., 2010), France (Girres and Touya, 2010) and Germany (Ludwig et al., 2011). Results indicate that populated places can reach higher positional accuracy and data completeness than the authoritative dataset, while less populated places are often characterised by lower accuracy and completeness. A challenge with the use of such traditional data quality assessment methods is that VGI dataset are now, in many parts of the world, more complete and accurate than authoritative datasets, violating the basic assumptions of those methods. To overcome these challenges, recent approaches explored the possibility to assess the quality of VGI data using intrinsic VGI properties such as the number of edits, the number of the contributors or a combination of them (Haklay et al., 2010; Mooney and Corcoran, 2012b; Keßler and Groot, 2013). By taking into account the evolving nature and the collaborative aspects of VGI, these approaches allow a better assessment of VGI data quality. Based on our literature review and on the work of Goodchild and Li. (2012) we summarize these newest quality evaluation methods in three main approaches.

2.1.1 Data-centric approach

Unlike traditional mapping activities, VGI result from collaborative mapping processes. To manage the collaborative aspect of maps' edition, most VGI projects use a version control system (VCS) that records all the changes (e.g., creation, edition, deletion) made to geographic objects. VCS allow retrieving and if necessary reverting back to a previous version of the objects. The data-centric approach takes advantage of this VCS to compute quality metrics.

Data-centric quality assessment methods were at first focused on the analysis of a single parameter. For example, Haklay et al. (2010) have analysed the relationship between the number of contributors and the positional accuracy of the data as objects in the database can be updated by many contributors. Results showed that high accuracy is achieved when there are at least 15 contributors per square kilometre. While positional accuracy is important in a map, having accurate and complete attributes is also of major importance for users. Mooney and Corcoran (2012a) have analysed how heavily edited object (i.e., more than 15 versions) are tagged in the OSM databases. As suggested by the authors, such heavily edited objects could gain the status of 'featured' objects and could be recognized for their quality.

Recent studies used a combination of indicators to provide a better estimation of the quality of VGI dataset. For example, Keßler and de Groot (2013) assessed the quality of OSM objects using a combination of five main parameters. Three of these parameters (i.e., number of versions, number of users and confirmations) have a positive influence on trustworthiness, while the two others (i.e., tags corrections and rollbacks) have a negative influence. Results from this analysis

showed that more than 75% of the selected features were correct when compared to ground-truthing data.

2.1.2 User-centric approach

Instead of deriving quality metrics from VGI data, the user-centric approach assesses the quality of VGI by evaluating contributors using different parameters such as contributors' motivations, experience or recognition (Rehrl et al., 2013). For example, if a contributor already has edits that have resulted in no or few modifications by other contributors in an active contributing area, this contributor could be identified as trustable. Such assessments could be easily applied to OSM, due to the fact that more than 80% of edits appear to be made by about 10% of contributors (Mooney and Corcoran, 2012b; Bégin and Devillers, 2013).

Several studies analysed the motivations of VGI contributors using different parameters, such as their goals, or their relations with other users (Coleman et al., 2009; Budhathoki et al., 2010; Neis and Zipf, 2012; Haklay, 2013). But these studies are often too generic to assess VGI data quality. One of the most interesting approaches is the one proposed by van Exel et al. (2011) who have coined the concept of 'crowd quality'. The crowd quality indicator is based both on an analysis of the users (e.g., experience, recognition) and on the data. Nevertheless, the relationship between the results of these analyses and VGI data quality has not yet been explored.

2.1.3 Context-centric approach

Studying the geographic context in which map elements are found can be used to enhance the quality of VGI data, although little work has been done using such an approach. For example, adding a pub in a place surrounded by nursing homes or in the middle of a park could be automatically identified as a potential error (Mülligann et al., 2011; Goodchild and Li, 2012). Such an approach relies on the knowledge of the context in which geographic information exist, being related to Tobler's first law of geography stating that "all things are related, but nearby things are more related than distant things" (Tobler, 1970). Nevertheless, this approach could require in practice an explicit schema that links the different geographical concepts. These links and the distance between concepts would allow measuring relationships between two sets of geographical concepts (e.g., pub and nursing homes) and possibly help identify potential errors.

In OpenStreetMap, two main projects have started exploring the possibility of creating a semantic schema of the OSM concepts by organizing them into an ontology. The first ontology, LinkedGeoData, aims to provide a geographic dataset for the semantic web. LinkedGeoData enriches OSM data with other semantic datasets such as DBPedia* or GeoNames** and then publishes them in a semantic form. At this date, LinkedGeoData contains more than 1 billion nodes, 100 million ways and approximately 20 billion triples (Auer et al., 2009). Despite the advantages of LinkedGeoData, this ontology is a simple tree structure where objects are linked using a relationships *is_a* (Ballatore et al., 2012). Moreover, due to the size of the OSM database, keeping a good scalability is still a major challenge (Auer et al., 2009). The second ontology, called OSMonto has been designed for use in a navigation web service (Codescu et al., 2011). Compared to LinkedGeoData, OSMonto provides a richer description of relationships between the OSM tags (e.g., *hasCuisine* relationships for a seafood restaurant). However, not all OSM concepts have been

* <http://dbpedia.org/>

** <http://www.geonames.org/>

implemented for performance and stability reasons. Moreover, as OSMonto has been designed for a specific application, reusing it in a generic context can be difficult. Despite the potential of these approaches, none has been officially integrated into the OSM infrastructure. In addition to these ontological approaches, two other projects led by Mülligann et al. (2011) and Ballatore et al. (2012) have been initiated. These projects use semantic similarity measurements to describe and evaluate relationships between OSM concepts. We describe these projects in more detail in the Section 3.2.

This section has highlighted the difficulties to assess the overall quality of VGI data and the need to test new quality evaluation methods. If many scientific studies have been done for the data-centric approach, the other two methods are still in their infancy. Moreover, these three approaches are not exclusive and can be combined to get a better assessment of VGI quality.

3. SEMANTIC SIMILARITY MEASUREMENTS

3.1 Semantic similarity concept and models

While humans can relatively easily discriminate between similar concepts, such a task is much harder to perform for computers. For example, if it seems natural for us to think of a 'road' as being more similar to a 'highway' than a 'river', creating the same knowledge in a computer system requires the use of specific computational models. One way to create and measure relationships between concepts is to use the concept of semantic similarity. Semantic similarity quantifies how similar, or dissimilar, concepts are based on their meanings. Several methods for determining semantic similarity between terms have been proposed in the literature. These methods can be classified into five main categories according to their knowledge representation and notion of similarity (Schwering, 2008): the geometric, feature, alignment, network and transformation models. A brief description of each of them is provided below. Detailed information about these methods and their applications in GIS can be found in Schwering (2008) and Li and Fonseca (2006).

- *Geometric model*: this method was first used in psychology. Semantic concepts are represented as a region in a multi-dimensional vector space (MDVS). Each dimension of this space is a property of the concept. The similarity between two concepts is represented as a vector distance between the two concepts in the MDVS.
- *Feature model*: each concept is defined by a collection of features. The semantic similarity between two concepts is expressed as a non-metric function of the common and distinct features of each set. While common features increase the similarity, distinct features decrease it.
- *Alignment model*: unlike the previous models, this model represents concepts in a structured way. The similarity between two concepts is computed by comparing common and different features, but also the relationships between these features.
- *Network models*: based on graph theory, concepts are modelled using semantic networks. In such networks, concepts are represented as nodes and relations between concepts are represented by edges. Like geometric models, the similarity between two concepts is measured using a distance function. This distance function is based on graph theory algorithms, such as the shortest path model.

- *Transformation model*: similarity is defined as a transformational distance that measures the number of transformations required to make one concept equal to another concept. In other words, fewer the transformations indicate more similar concepts.

These five main categories are not exclusive and hybrid approaches can be used to measure the semantic similarity between two concepts.

3.2 Semantic similarity applied to VGI

In OpenStreetMap, real-world phenomena are described by associating a set of attributes to geographic primitives. In the current OSM application programming interface (API 0.6*), four types of geometric primitives (point, line, polygon and relation) can be created. Attributes are described using structured key/value pairs called 'tags', where the 'key' is similar to an attribute and 'value' is the value of this attribute for the given geometric object. For example a university can be described using a point or a polygon with the associated tag "*amenity=university*". An OSM object can be tagged with any number of tags but each key for a given geographic object has to be unique. These tags are the result of informal and continuous discussions within the community and the most popular tags are described in the OSM wiki**. However there is no restriction on the use of tags, and contributors can freely create their own tags when necessary (Mooney and Corcoran, 2012a). While such flexible community-based approach allows for a rich semantic, it also creates a significant unnecessary semantic heterogeneity as contributors can be tempted to use rare tags or to create their own tags when they do not know the existence of a tag commonly used by the community to describe an object. Such flexibility can also result in a lack of consensus on tags to use. Some OSM users will for example tag a forest as "*landuse=forest*" while others prefer "*natural=wood*" (Ballatore et al., 2012). Two main projects have looked at the problem of semantics in VGI.

The first project conducted by Mülligann et al. (2011) combined semantic similarities with a point pattern analysis in a two step method to study the geography of OSM semantics. First, the semantic similarity was computed from the OSM dataset by looking at the spatial co-occurrence of features. Second, spatial-semantic pattern were identified indicating potential correlations between two different types of geographic objects or between two same kinds of geographic objects. For example, the addition of a new fire station close to an existing one (i.e., average distance between these two features below the spatial-semantic pattern of this feature) could be detected as a potential error.

The second project proposed by Ballatore et al. (2012) used a semantic network created from the extraction of the relationships between OSM concepts in the wiki. This extraction is performed using the OSM Wiki Crawler developed by the authors. In this semantic network, OSM concepts are represented as vertices and edges between these concepts are computed by measuring the relatedness between the wiki pages. Using this semantic network, different similarity measures (e.g., *SimRank*, *P-Rank*) based on the co-citation algorithm are tested and a semantic similarity score between two concepts was computed.

* http://wiki.openstreetmap.org/wiki/API_v0.6

** http://wiki.openstreetmap.org/wiki/Map_Features

4. PROTOTYPE DESCRIPTION AND EXAMPLES

4.1 OSM Semantic Plugin

This research integrates semantic similarity measurements into a VGI editor to support users during the data editing process. This section presents the semantic plugin and how semantic similarity is computed.

4.1.1 Prototype description

The most commonly used OSM*** editor is the Java OpenStreetMap (JOSM) Editor. JOSM is a free and Open Source stand-alone desktop application that allows contributors to create, edit or delete data from OSM. JOSM has a flexible API that allows creating plugins that extend JOSM basic functionalities. JOSM was used for its popularity and flexible API.

We developed a new plugin called “*OSM Semantic Plugin*” to improve the intrinsic semantic heterogeneity of OpenStreetMap data (Figure 1). The plugin measures semantic similarities between OSM tags to automatically suggest other similar tags and to display a warning when a too high or a too low similarity value has been detected. Once the plugin is installed, a new window called *OSMantic* is added on the right panel of JOSM and users can freely move and resize the window to fit to their needs.

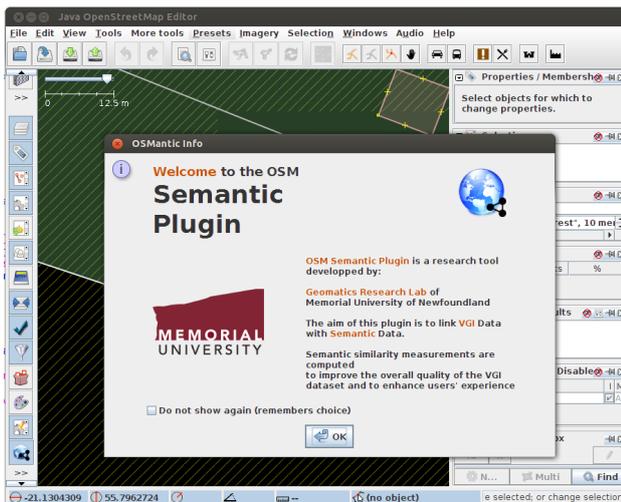


Figure 1. Integration of the *OSM Semantic Plugin* in JOSM

Because OSM is a global mapping project, we recognize that people from different countries or regions may have differences in the way they want to conceptualize geographic phenomena. To support such important semantic diversity, while reducing unnecessary local semantic heterogeneity, we based our approach on Tobler's first law of geography (Tobler, 1970) by viewing semantic homogeneity as being higher for objects that are closer in space. In other words, contributions closer in space are more likely to use similar semantic to describe objects than contributions that are more distant. The *OSM Semantic Plugin* implements this approach by analysing the semantics of existing objects located in the neighbourhood of the object being created.

4.1.2 Semantic Similarity Computation

The OpenStreetMap semantic network used to measure the semantic similarity between concepts has been created using

*** http://wiki.openstreetmap.org/wiki/Editor_usage_stats

the OSM Wiki Crawler from Ballatore et al. (2012). The OSM Wiki Crawler extracts OSM tags and their relationships (hyperlink) from the OSM wiki website and computes the semantic similarity score of each pair of tags. The P-Rank (Penetrating Rank) co-citation algorithm (Zhao et al., 2009) has been chosen to compute the semantic similarity score between sets of tags. Compared to other algorithms (e.g., SimRank) that only take into account incoming links, P-Rank is a recursive algorithm based on both incoming and outgoing links of the networks. In other words, in P-Rank: “two entities are similar if (1) they are referenced by similar entities and (2) they reference similar entities” (Zhao et al., 2009). In P-Rank, the semantic similarity denoted $s(a, b) \in [0, 1]$ between two vertices (a, b) of a network G where $a, b \in G$ is defined as (Equation 1):

$$R_{k+1}(a, b) = \lambda \times \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} R_k(I_i(a), I_j(b)) + (1 - \lambda) \times \frac{C}{|O(a)||O(b)|} \sum_{i=1}^{|O(a)|} \sum_{j=1}^{|O(b)|} R_k(O_i(a), O_j(b)) \quad (1)$$

Equation 1. The P-Rank formula (Zhao et al., 2009)

Where $R_k(a, b)$ is the P-Rank similarity score between two vertices a and b on iteration k and assuming that $R_k(a, b)=1$ if $a=b$ and $a \neq b$. The P-Rank similarity score is updated at each iteration until the maximum number of iterations (k) is reached. At the end of the iterations, the P-Rank similarity score is a value between zero (i.e., no semantic similarity) and one (i.e., tags are similar).

The *OSM Semantic Plugin* currently works for points, lines and polygons OSM primitives. Further work will have to be done to handle OSM relation primitives that are more conceptual than geographic objects.

4.2 Examples

4.2.1 Creation of a new OpenStreetMap object

When a new OSM object is created, geographic objects in its neighbourhood are analysed to automatically propose similar or relevant tags. For example, if a user creates or edits an OSM object located in a city, related tags such as restaurant or shop, will automatically be proposed. In the current version of the plugin, the search radius that determines the size of the neighbourhood is fixed and cannot be changed but future work will allow users to set their radius and explore how an optimal radius can be defined.

Figure 2 illustrates the park example described above. After the creation of a new point (the red point on Figure 2), the *OSM Semantic Plugin* analysed all the tags in the neighbourhood to automatically suggest relevant tags. As the new point is located close to several restaurants, the semantic plugin has automatically suggested OSM concepts related to commerce, such as the tag “shop=*general_store*” or the key “diet:vegan”.

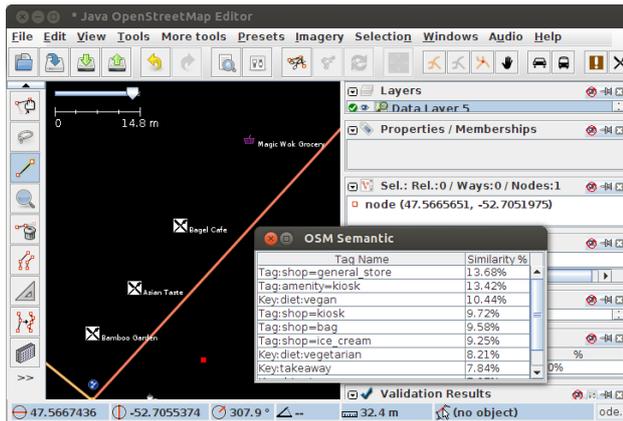


Figure 2. Tags and keys proposed by the OSM Semantic Plugin during the creation of a new object (i.e., red point)

All similar tags and their similarity score are displayed in a table (cf. Figure 2). The semantic similarity score, initially between 0 to 1, was converted to a percentage to be easier for people to understand. When a contributor selects a tag from the results table, this tag is automatically added as a property of the new OSM object, also encouraging a richer description of objects.

4.2.2 Edition of an existing OpenStreetMap object

Precisely describing an OSM object often requires using several tags. However, the current OSM schema contains more than 40,000 tags and contributors cannot know them all and currently have no guidance on the best use of those tags. Improving the contribution process by automatically suggesting related tags can help the contributor to provide a better description of OSM objects. When editing an object that already has a set of tags, the plugin takes the tags into account and combines them with the OSM tags from similar objects around it. For example, a traffic signal can be generally described by the tag “highway=traffic_signals”. However, other tags could be added to extend the description of the object, such as a tag “traffic_signals:sound=yes/no” for traffic signals equipped with sound that supports visually impaired people (Figure 3).

Figure 3 illustrates this case. If a contributor selects an existing OSM object tagged with traffic signal attributes, similar keys such as *traffic_signals* with *sound* or *floor_vibration* values are automatically proposed.

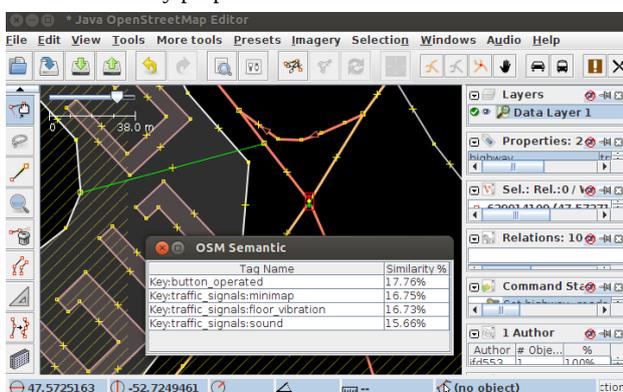


Figure 3. Suggestion of similar tags for an existing object

4.2.3 Reducing semantic heterogeneity

Two types of warnings were implemented to prevent or reduce semantic heterogeneity during the editing process.

The first type prevents from adding tags that are too dissimilar. When a user adds a new set of tags to an OSM object, the semantic similarity between these new tags and the existing ones is computed. If the semantic similarity score is lower than 8% (i.e., an arbitrary threshold that can be changed), then a warning is automatically displayed to the user. Figure 4 illustrates such a case with a contributor editing an object initially tagged as “leisure=playground” by adding a new tag “highway=roads”. As the similarity between these two tags is low, a warning is displayed in the system.

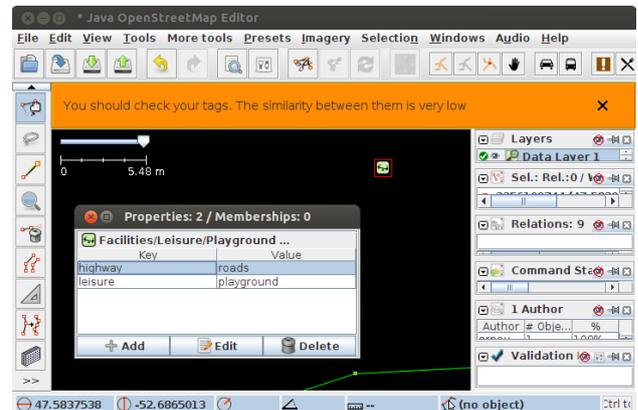


Figure 4. Preventing semantic heterogeneity by warning the contributor

The second type of warning aims to reduce the overall semantic heterogeneity amongst objects located in a same area. This analysis takes into account all the objects located in the neighbourhood of the newly created object and displays a warning when a high similarity score is detected. For example, if a user wants to describe a forest using the tag “landuse=forest” while all the other OSM objects in the neighbourhood are tagged with “natural=wood”, then a warning is automatically displayed.

Using the OSM Semantic Plugin can enhance the quality of the contributions and reduce the semantic heterogeneity of the VGI data, improving the quality of the dataset. While this plugin has been developed and tested, it is not yet available in the JOSM plugin repository as improvements have to be done before releasing it publicly.

5. CONCLUSIONS

The growing popularity of VGI projects such as OpenStreetMap changes traditional geographic information production, dissemination and use. While such free and up-to-date data can often be a valuable alternative to authoritative datasets, assessing and enhancing the quality of VGI dataset remains a significant challenge. Specifically, the high semantic heterogeneity of VGI datasets can be a major obstacle to the use of such data for diverse data analysis.

This paper has investigated the use of semantic similarity to enhance the quality of VGI data by reducing semantic heterogeneity in a given neighbourhood. To test our approach, a plugin integrated into the JOSM editor was developed and three examples of its use were presented. These examples have illustrated potential situations where the use of semantic similarity measurements could reduce the semantic heterogeneity.

Further work will allow improving the approach. First, the OSM Semantic Plugin could be extended to handle the recent OSM primitive ‘Relation’ that is increasingly popular amongst contributors. Second, semantic similarity results could be more

accurate by computing not only the similarity between tags, but also the similarity between sets of tags that describe a geographic object.

Acknowledgements

This research was funded by the Natural Science and Engineering Research Council of Canada (NSERC). Authors also thank Andrea Ballatore for sharing his results and Vincent Privat for his helpful feedback on JOSM.

6. REFERENCES

- Auer, S., Lehmann, J., Hellmann, S., 2009. LinkedGeoData: Adding a Spatial Dimension to the Web of Data. Proceedings of the 8th International Semantic Web Conference. Springer-Verlag, Berlin, Heidelberg, pp. 731–46.
- Ballatore, A., Bertolotto, M., Wilson, D.C., 2012. Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems (KAIS)*, pp. 1–21.
- Bégin, D., Devillers, R., Roche, S., 2013. Assessing Volunteered Geographic Information (VGI) quality based on contributors' mapping behaviours. Proceedings of the 8th International Symposium on Spatial Data Quality. China, Hong Kong.
- Budhathoki, N.R., Nedović- Budić, Z., Bertram, B., 2010. An Interdisciplinary Frame for Understanding Volunteered Geographic Information. *The Journal of Geospatial Information Science, Technology and Practice*, 64(1), pp. 11–26.
- Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., Rau, R., 2011. *DO-ROAM: Activity-Oriented Search and Navigation with OpenStreetMap*. GeoSpatial Semantics. Springer Berlin Heidelberg, pp. 88–107.
- Coleman, D.J., Georgiadou, Y., Labonte, J., 2009. Volunteered Geographic Information: the nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4, pp. 332–58.
- Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., Shi, W., 2010. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. *Transactions in GIS*, 14(4), pp. 387–400.
- Elwood, S., Goodchild, M.F., Sui, D.Z., 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3), pp. 571–90.
- Flanagin, A.J., Metzger, M.J., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), pp. 137–48.
- Girres, J.-F., Touya, G., 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14(4), pp. 435–59.
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211–21.
- Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, pp. 110–20.
- Haklay, M., 2013. *Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation*. Crowdsourcing Geographic Knowledge. Springer Netherlands, pp. 105–22.
- Haklay, M., Basiouka, S., Antoniou, V., Ather, A., 2010. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4), pp. 315–22.
- ISO/TC 211, 2002. 19113 Geographic information – Quality principles, International Organization for Standardization. International Organization for Standardization (ISO).
- Keßler, C., de Groot, R.T.A., 2013. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap. Association of Geographic Information Laboratories for Europe. Leuven, Belgium, pp. 16.
- Li, B., Fonseca, F., 2006. Tdd: A comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation*, 6, pp. 31–62.
- Ludwig, I., Voss, A., Krause-Traudes, M., 2011. *A Comparison of the Street Networks of Navteq and OSM in Germany*. Advancing Geoinformation Science for a Changing World. Springer Berlin Heidelberg, pp. 65–84.
- Mooney, P., Corcoran, P., 2012a. The Annotation Process in OpenStreetMap. *Transactions in GIS*, 16(4), pp. 561–79.
- Mooney, P., Corcoran, P., 2012b. Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, 4(4), pp. 285–305.
- Mülligann, C., Janowicz, K., Ye, M., Lee, W.-C., 2011. Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. Proceedings of the 10th international conference on Spatial information theory. Springer-Verlag, Berlin, Heidelberg, pp. 350–70.
- Neis, P., Zipf, A., 2012. Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(3), pp. 146–65.
- Rehrl, K., Gröechenig, S., Hochmair, H., Leitinger, S., Steinmann, R., Wagner, A., 2013. *A Conceptual Model for Analyzing Contribution Patterns in the Context of VGI*. Progress in Location-Based Services. Springer Berlin Heidelberg, pp. 373–88.
- Schwering, A., 2008. Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. *Transactions in GIS*, 12(1), pp. 5–29.
- Tobler, W., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, pp. 234–40.
- Van Exel, M., Dias, E., 2011. Towards a Methodology For Trust Stratification in VGI. Volunteered Geographic Information (VGI) – Research progress and new developments. Seattle, Washington, pp. 4.
- Zhao, P., Han, J., Sun, Y., 2009. PRank: a comprehensive structural similarity measure over information networks. Proceedings of the 18th ACM Conference in Information and Knowledge Management. ACM Press, pp. 553–62.