YOU DESCRIBE IT, I WILL NAME IT: AN APPROACH TO ALLEVIATE THE EFFECT OF USERS' SEMANTICS IN ASSIGNING TAGS TO FEATURES IN VGI

S. Hassany Pazoky, F. Karimipour, F. Hakimpour

Faculty of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, North Kargar St., Tehran, Iran
(shpazooky, fkarimipr, fhakimpour)@ut.ac.ir

Commission II, WG II/4

KEY WORDS: Uncertainty, VGI, Users' Semantics, Tagging Features, OpenStreetMap, Perception

ABSTRACT:

As an important factor of VGI quality, this paper focuses on uncertainty arisen in assigning tags to features by VGI users. The VGI portals ask their users to assign (or tag) one or more data types to features, from a set of pre-defined types, whose meanings may be vague for the user, or distinctions between some of them are not clear, i.e. depend on the users' semantics. This research believes such uncertainties are the results of perceptual issues arising in serial communication between the system and the user. We categorize the problem, and then utilize semantic modelling to reduce such uncertainties. A hierarchy of feature types is produced. At each step, users are asked a simple question with clear distinct answers, which gradually directs the user to the right type. We will describe the approach and present the initial results for the hierarchy produced for major linear features of OpenStreetMap.

1. INTRODUCTION

Among the technological advances of the third millennium, Web 2.0 in line with themes such as crowdsourcing, collaboration, wikis, and the GeoWeb have thoroughly changed the World Wide Web. Neogeography was a pioneer term introduced by Turner (2006) to convey the idea of participation of untrained users in map production process. In other words, the distinction between map producer, communicator and user loses clarity (Goodchild, 2009b). Goodchild (2007a) took advantage of this research and proposed a new term namely Volunteered Geographic Information (VGI), which is widely used in the literature and various research are being conducted on different aspects of it.

One of the most important challenges of VGI development is uncertainty mentioned by many researchers who have considered it from different aspects (Allingham, 2014; Barron et al, 2013; Elwood et al, 2013; Mohammadi and Malek, 2015). The research conducted on VGI uncertainty have mostly measured the uncertainty of VGI datasets (Vandecasteele, and Devillers, 2015), 2) or concerned with spatial aspect of the data rather than non-spatial aspect (Mülligann et al., 2011; Mooney and Corcoran, 2012). To the best of our knowledge, there are only a very few attempts to introduce mechanisms to lower the uncertainty during the editing process (Grira et al., 2010; Vandecasteele, and Devillers, 2015).

This paper focuses on an approach to reduce uncertainty arisen in assigning tags to features by VGI users during the editing process. The idea is based upon the fact that citizens, who play the role of surveyors in VGI, provide the required non-spatial information based on their perception (Flanagin and Metzger, 2008; Mülligann et al., 2011; Karimipour et al., 2013; Mount, 2013), social and cultural settings (Goodchild, 2007a; Goodchild, 2007b; Coleman et al, 2009; Roche et al., 2012), previous experience (Mülligann et al., 2011), etc. In other words, substituting neo-geographers with experts is equivalent to substituting perception with measurements, or quality with quantity. In the case of experts, the hidden information is about the measuring instruments where there are many standards to

reduce them under an acceptable threshold. As a result, voluntarily generated maps are more suitable for everyday life and recreation, as they are more up-to-date, whereas maps produced by experts are more suitable for engineering purposes.

OpenStreetMap asks contributors to assign (or tag) at least one data type to features, whose meanings may be vague for the user, or distinctions between some of them are not clear, i.e. depend on the users' semantics. As a result, unlike spatial information, data type information are accompanied with hidden semantic. In other words, different people may tag features differently leading to semantic heterogeneity (Vandecasteele, and Devillers, 2015).

Mülligann et al. (2011) is one of the first efforts that concern on the problem of tagging features in VGI. They developed a semantic similarity measure to assist contributors in tagging point features. They use spatial relationships to outline incorrect tags. For example, a pub is usually surrounded with places that can afford drinking alcohol, while waste baskets are distributed uniformly in the city. As another example, tagging two very near features as "fire stations" may be an indicator of a mistake by the contributors.

Ballatore et al. (2013) also developed a semantic network by crawling OSM wiki page to measure the semantic similarity between feature types. The outcome of the paper can be used for recommender systems to tag features in OpenStreetMap, geographic information retrieval, and data mining.

Vandecasteele and Devillers (2015) used the semantic network produced by Ballatore et al. (2013) and combined it with TagInfo to form their database. They developed an open source plugin called OSMantic to recommend tags to users and also warn them when inappropriate tags are used together.

This paper proposes the initial result of a solution for the problem of assigning tags to features in VGI. We believe such uncertainties are the results of perceptual issues arising in serial communication between the system and the user. We categorize the problem, and then utilize semantic modelling to reduce such uncertainties. A hierarchy of feature types is produced. At each step, users are asked a simple question with distinct enough answers, which gradually directs the user to the right type. In

addition, users are not forced to tag the features that they really do not have information about. They can proceed in the hierarchy as much as they have information, i.e. as much as they can answer the questions.

The rest of the paper is structured as follows: In Section 2, we scrutinize the perceptual causes of uncertainty that bothers VGI, especially OpenStreetMap. Section 3 describes the proposed approach. Sections 4 presents the initial results of deploying the proposed approach to produce the hierarchy for major linear features of OpenStreetMap, as a successful VGI portal (Haklay & Weber, 2008; Ballatore et al., 2013). Finally Section 5 concludes the paper.

2. PERCEPTUAL ISSUES OF SERIAL COMMUNICATION

Communications may be considered either as parallel or serial. The former makes use of pictorial elements that are perceived all together. On the other hand, means of communication in serial communication are words that are perceived one by one and in a predefined order.

Whenever serial communication is used, measures should be devised to counter the uncertainty. If a neogeographer is asked to digitize a street for a VGI portal, he most probably picks the right drawing tool and draws a simple line. However, the problem arises when he is to tag it with the appropriate term. The perceptual issues of tagging features in VGI are as follows:

(1) What do the VGI portal administrators mean by a certain tag such as "highway" tag? Such terms are among the professional contexts. There are many institutes in the world defining these terms. The worst aspect is that they may have different meanings in different countries or they may have special local meaning in some places. For example, most natural water areas in Newfoundland, Canada are named "ponds". However, in OpenStreetMap, "pond" refers to areas of water created by human activity[1] (Vandecasteele & Devillers, 2015). Even, some countries may have other feature types rather than those available in OpenStreetMap.

(2) A major aspect of well-known VGI portals such as OpenStreetMap is that they support multi-linguality, i.e. the tags are provided in many different languages. In addition to emphasizing our concerns on translating tags in OpenStreetMap, the tags in countries speaking the same language may even differ. For example, the tag "highway=motorway", which is used in UK, is equivalent to "motorway, freeway, and freeway-like road" in Australia, "limited access highway" in Canada, "freeway" in India, and "limited access freeway" in USA[2]!

(3) There are too much tags in OpenStreetMap. The number of tags in OpenStreetMap is much more than a human can perceive in mind for voluntary actions. In addition, users can create their own tags (Mooney and Corcoran, 2012). This is why there are many miss-spelled tags in OpenStreetMap data. Vandecasteele and Devillers (2015) report that there are currently more than 40,000 distinct tags in OpenStreetMap data which is tens of times more than what is really needed.

(4) The terms used to assign tags to features are not clear enough for users (Ballatore et al, 2013). For example, at the first glance, the tags such as "highway=primary", "highway=secondary", "highway=tertiary", "highway=residential", and "highway=living_street" may seem indistinguishable.

Especially, in ancient cities where urban planning has had less opportunity to develop the city in an organized manner, the differentiation of the aforementioned types is a professional task, if not impossible. This problem is regarded as ambiguity (Shi, 2010). On the other hand, contributors of VGI portals are ordinary people without special training in GIS or similar fields. Thus, they tag map features based on their perception or semantic. Naturally, users' semantics are not necessarily the same, which can make VGI imbalanced.

(5) Although OpenStreetMap has done its best to provide clear definitions of different feature types, the extent of some of which may not be clear enough. For example, "highway=tertiary" tag is defined in OpenStreetMap as a tag "used for roads connecting smaller settlements, and within large settlements for roads connecting local centres. In terms of the transportation network, OpenStreetMap tertiary roads commonly also connect minor streets to more major roads.[3]" And, residential road is defined as: "Roads accessing or around residential areas but are not a classified or unclassified highway." It is then mentioned that if you doubt whether the road is residential or unclassified, residential is more specifically defined as: "Street or road generally used only by people that live on that road or roads that branch off it."[4] Although the two definitions are in clear English, they do not share a clear boundary. For example, there are paragraphs describing "highway=tertiary" and "highway=residential_road" tags; but they do not share a clear boundary; i.e. there are many instances that apply to both. As a result, contributors may face problems instantiating real world objects and they use the tags interchangeably depending on their semantics (Vandecasteele & Devillers, 2015). This phenomenon is regarded as vagueness by Shi (2010).

(6) The contributors may lack information that causes imprecision (Shi, 2010). If so, they are unable to provide the portal with the right tag. An illustrative example of imprecision is the location of Eiffel tower. Although answers such as Europe, France, and Paris are all correct with no uncertainty, Europe is the less and Paris is the most precise notion.

Also, Vandecasteele and Devillers (2015) include temporal changes of tags as a source of semantic heterogeneity. We do not agree with them since tags that are not used become deprecated and they can easily be distinguished and replaced with the new tags.

The sextet uncertainty elements caused by users' semantics defined above occur when contributors wish to tag a feature. The effect of them, however, is not equal and it may vary from a person to another. Semantic can help resolve the perceptual problems.

3. HOW TO ALLIVIATE THE EFFECT OF USERS' SEMANTICS?

This paper proposes a solution to alleviate the effect of users' semantics in assigning tags to features in VGI. The solution provides clear definitions of different object types dragging the understanding of the system designers and users together as much as possible. We extract the information of the user, purify it from any irrelevant semantics and use it in VGI. For this, all the possible tags are arranged in a hierarchical structure. The contributor goes through the hierarchy answering some questions asked at each node. The questions are very simple, free of

---

[1] http://wiki.openstreetmap.org/wiki/Tag:landuse%3Dpond

[2] http://wiki.openstreetmap.org/wiki/Highway:International _equivalence

[3] http://wiki.openstreetmap.org/wiki/Tag:highway%3Dtertiary

[4] http://wiki.openstreetmap.org/wiki/Tag:highway%3residential

complex and technical terms, and qualitative. Also, the choices available for every questions are also simple and completely distinct. The benefit of the hierarchical structure is that at each node, the contributor faces a clear question with very few possible answers; whereas in the current approach of OpenStreetMap, the variety of choices without clear distinctions confuses the contributor. This procedure gradually directs the contributor to the right data type. In addition, the contributor can stop descending the hierarchy when she lacks information. This mechanism provides the opportunity to get information from the contributors to the extent they are sure about, i.e., neither more nor less than what they really know!

4. CASE STUDY: TAGGING LINEAR FEATURES IN OSM

To illustrate the proposed idea of the paper, which is also briefly presented in Pazoky et al. (2014), this section describes how the idea is applied on the linear features available at the official wiki page of OpenStreetMap[5]. We produced the hierarchy for aeroway, highway, railway, and waterway types from the 26

available feature types. 38 tags were chosen among these feature types for the case study.

To design the questions and hierarchy, we started the hierarchy with OpenStreetMap categorization of feature types mentioned before. The rest of the hierarchy was produced using a divisive approach, i.e. a member of a category was chosen and its most significant distinction was considered to form the question. However, if the question in its parent node was less general, the two questions were swapped. Then, all the elements were reconsidered and moved if they belong to other nodes.

The resultant hierarchy for the mentioned data types is illustrated in Figure 1. The light green boxes show the provided choices; and the darks green boxes indicate the OpenStreetMap tag associated with that feature type. Furthermore, the blues boxes in Figure 1 are expanded separately in Figures 2 to 5 for "Cars and pedestrians", "Link between two roads", "Rail network", and "water" respectively. The path taken by a hypothetical contributor is shown in Figure 6.
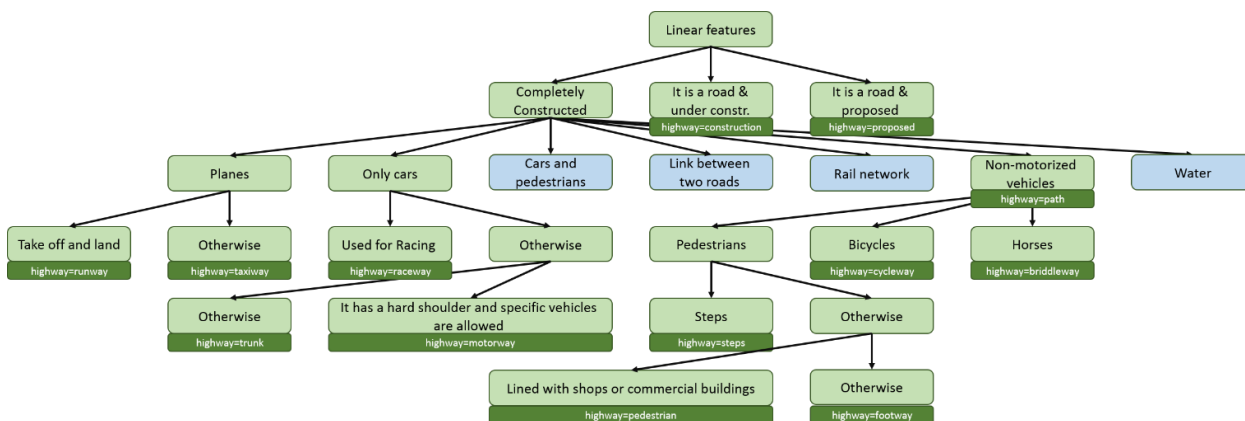

Figure 1. The resultant hierarchy to tag OpenStreetMap features.

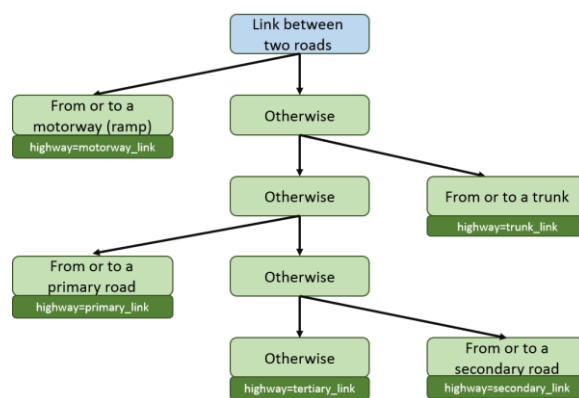Figure 2. Expanded view of "Cars and pedestrians" box in Figure 1.

Figure 3. Expanded view of "Link between to roads" box in Figure 1.
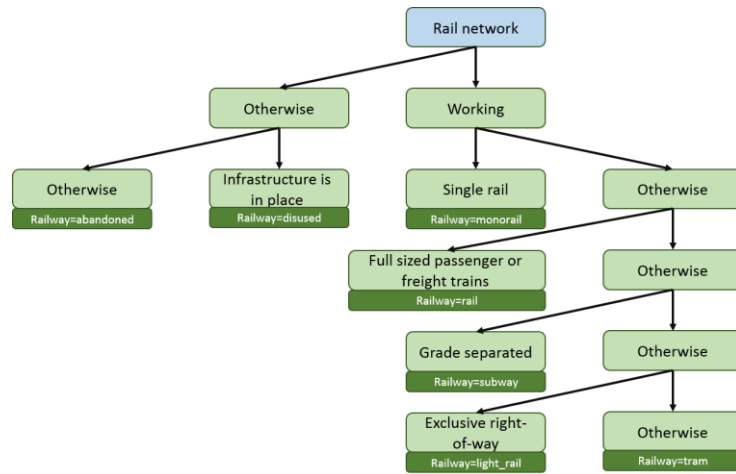
Figure 4. Expanded view of "Rail network" box in Figure 1.
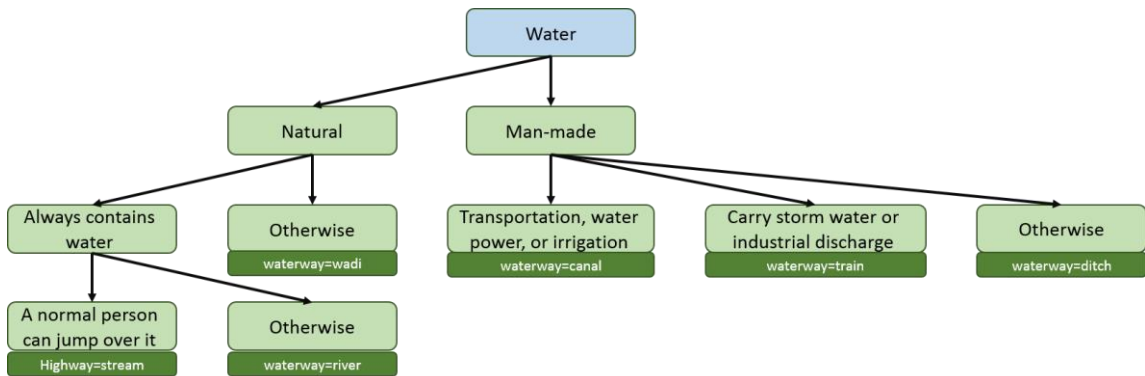

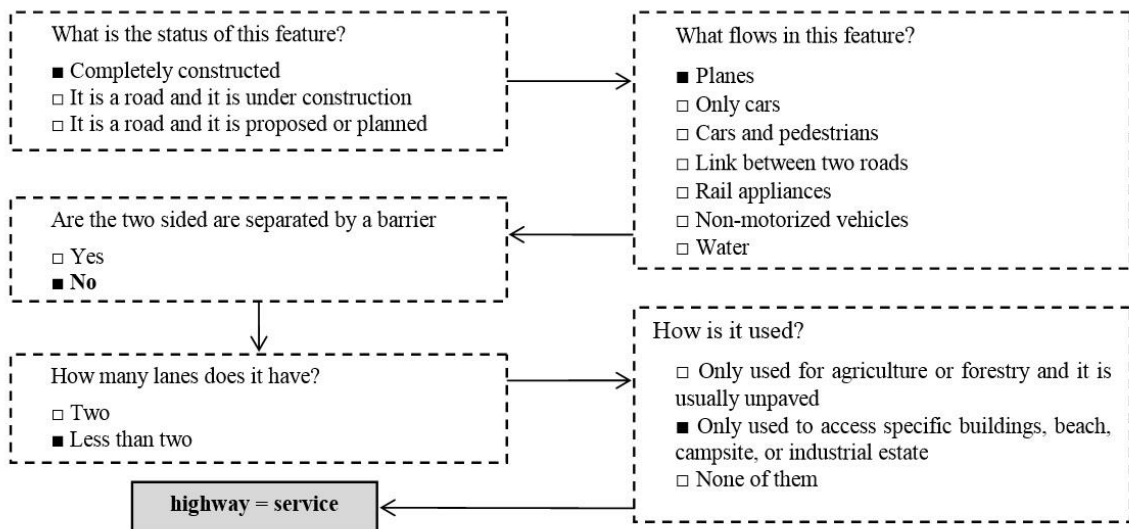Figure 5. Expanded view of "Water" box in Figure 1.

Figure 6. The questions and answers our hypothetical contributor has gone through to reach the highway=service tag.

5. CONCLUSION AND FUTURE WORK

VGI, as a prominent innovation of the past decade in GIScience (Goodchild, 2009a), has been very fruitful in gathering geospatial information from the general public. In this paper, we introduced users' semantics or differences between the perceptions of people on a term as a significant source of uncertainty in VGI. This problem is caused by using complex and professional terms in VGI portals such as OpenStreetMap to tag features. We proposed a solution to alleviate the effect of users' semantics on the issue. Using the hierarchy provided in the solution, contributors are faced with a sequence of questions answering of which leads them to the next questions until they reach the leaf. This way, professional terms are avoided and a thick barrier between the possible answers is established, which result in clear distinct choices. The case study showed the professional terms with vague boundaries are reachable through the hierarchy. This solution lets the VGI portals receive less uncertain information from untrained and inexperienced contributors. It can also be regarded as a step to develop ways to diminish perceptual uncertainties of VGI.

We are going to develop an application implementing our approach to be testable by different people. Then, we intend to ask people from around the world to tag map objects using our approach and see if it really works and gather their feedbacks to improve the usability, user-friendliness, misclassifications, etc. Another goal of us is to expand the hierarchy to accommodate all the feature types of OpenStreetMap, i.e. point, linear, and polygonal objects.

REFERENCES

Allingham, C. (2014). A review of quality of geo-data from user's perspective. Universal Journal of Geoscience 2(2), pp. 70-75. DOI=10.13189/ujg.2014.020205.

Ballatore, A., Bertolotto, M., & Wilson, D. C. (2013). Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems*, *37*(1), 61-81.

Barron, C., Neis, P., and Zipf, A. (2013). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*. DOI=10.1111/tgis.12073.

Coleman, D. J., Georgiadou, Y., and Labonte, J. (2009). Volunteered geographic information: the nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, *4*(1), pp. 332-358.

Elwood, S., Gooodchild, M. F., and Sui, D. (2013). Prospects of VGI research and the emerging fourth paradigm. In: *Crowdsourcing Geographic Knowledge,* Springer Netherlands, pp. 361-375. DOI= 10.1007/978-94-007-4587-2_20.

Flanagin, A. J., and Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137-148.

Goodchild, M. F. (2007a). Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4), pp. 211-221. DOI=10.1007/s10708-007-9111-y.

Goodchild, M. F. (2007b). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructure Research*, 2, pp. 24-32.

Goodchild, M. F. (2009a). Geographic information systems and science: today and tomorrow. *Annals of GIS*, 15(1), pp. 3-9. DOI=10.1080/19475680903250715.

Goodchild, M. F. (2009b). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, 3(2), pp. 82-96. DOI=10.1080/ 17489720902950374.

Grira, J., Bédard, Y., and Roche, S. (2010). Spatial data uncertainty in the VGI world: going from consumer to producer. *Geomatica*, 64(1), pp. 61-71.

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps.*Pervasive Computing, IEEE*, 7(4), 12-18.

Karimipour, F., Esmaeili, R., and Navratil, G. (2013). Cartographic representation of spatial data quality parameters in volunteered geographic information, *Proceedings of the 26th International Cartographic Conference (ICC 2013),* (Dresden, Germany, August 25-30, 2013).

Mohammadi, N., & Malek, M. (2015). Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. *Journal of Spatial Science*, *60*(1), 119-135.

Mount, J. (2013). The role of context in spatial decision-making in GIScience. *Ph.D. Dissertation, University of Iowa.*

Mülligann, C., Janowicz, K., Ye, M., & Lee, W. C. (2011). Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. In *Spatial information theory* (pp. 350-370). Springer Berlin Heidelberg.

Mooney, P., & Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS*, *16*(4), 561-579.

Pazoky, S., Karimipour, F., and Hakimpour, F. (2014). An Ontological Solution for Perceptual Uncertainties of VGI. In: *Proceedings of the Extended Abstracts of 8th International Conference on Geographic Information Science (GIScience 2014)*, Vienna, Austria, 23-26 September 2014, 302-305.

Roche, S., Mericskay, B., Batita, W., Bach, M., and Rondeau, M. (2012). Wikigis basic concepts: Web 2.0 for geospatial collaboration. *Future Internet*, 4(1), pp. 265-284.

Shi, W. (2010). Principles of modeling uncertainties in spatial data and spatial analyses. CRC Press.

Vandecasteele, A., & Devillers, R. (2015). Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In *OpenStreetMap in GIScience* (pp. 59-80). Springer International Publishing.