ONTOLOGICAL ASSESSMENT AND SIGNIFICANCE OF SEMANTIC LEVELS OF TAGS IN OPENSTREETMAP

S. Ahmadian [1,2]*, F. Hakimpour [1]

[1]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2] Mazandaran Regional Electric Company
(s_ahmadian, fhakimpour@ut.ac.ir)

KEY WORDS: Volunteered Geographic Information, semantic heterogeneity, semantic levels, Rough Set Theory, significance, OpenStreetMap

ABSTRACT:

User-generated contents are developing rapidly through VGI and contributors create the tags through the Web applications in a free mechanism. Semantic Knowledge in VGI like other user-generated contents needs to be combined with other authoritative data sources. One of the main challenges of integration is the semantic heterogeneity of user-generated contents which are describing the geographical objects as POIs. Geographical objects can be described in different semantic levels such as purpose or function. Significance of semantic levels defines the importance of related attributes. Analysis of significance for semantic levels of different POIs can be considered as a base to enhance the semantic quality of VGI. This paper proposes an approach based on the notions of rough set theory to measure the significance of semantic levels of tags which are applied to describe the buildings in OpenStreetMap. The proposed approach is implemented for tags which are applied to describe buildings in OpenStreetMap. Results show the high significance for tags which describing the semantic levels of geographic information constructs and purpose/ function for buildings.

1. INTRODUCTION

1.1 Semantic heterogeneity in VGI

The emergence of participative web and pervasive internet by portable location-enabled devices result in creating a huge amount of volunteered geographic information (VGI). Many types of VGI are developed in forms of semi-structured text, image and vector data(Senaratne, Mobasheri, Ali, Capineri, & Haklay, 2016).

OpenStreetMap is a global effort to create a detailed global map using volunteers' effort and has evolved to be one of the greatest and most famous VGI project. Tags in OpenStreetMap consist of a pair of key and value. For example, according to the categories of Map Features, contributors are recommended to describe the universities by "Amenity=University" or "Building=University". Contributors usually are not aware of the semanticinformation of geographical objectsor hierarchical relations. Besides, there is no or a little knowledge about explicit or implicit semantic relations between geospatial objects and finally, tags are not prioritized to show the significance of each tag for different types of geospatial objects. Based on (Girres & Touya, 2010; Senaratne et al., 2016), standardized specification and classification will improve the semantic accuracy of OpenStreetMap data. Free mechanism of contribution and different semantic levels of describing geographical objects as Points of Interest (POIs) causes to semantic interoperability problems. In the case of VGI, different conceptual perception of POIs is one of the sources of semantic heterogeneity. Vandecasteele and Devillers (2015)introduced main sources of semantic heterogeneity in VGI as nature of the concept, the geographic scale a concept is used at and finally temporal evolution of concept definition

According to the works of Bishr and Kuhn(2007), semantic heterogeneity is one of the main problems of collaboratively generated geospatial contents. Geo–ontologies as the semantic frameworks provide tools to resolve the semantic problems such as lack of hierarchy, synonym control, and semantic precision between different geospatial data sources.

Different ontologies are studying the way of formal representation of geographical objects (Ballatore, 2016; Frank, 2001; Kuhn, 2003; Scheider& Janowicz, 2014; Scheider, Janowicz, & Kuhn, 2009; Smith & Mark, 2001). Ontologies such as DOLCE and WordNet apply cognitive and linguistic perspective focusing on the semantics of natural language terms and the ways of structuring human understandings of space. Many approaches to spatial cognition are presented in the work of cognitive linguists (Bishr& Kuhn, 2007; Couclelis, 2010; Kuhn, 2003). One of the convenient geo-ontologies to study the way of conceptualization of geographical objects is the semantic levels of Couclelis (2010). She introduced the main semantic levels for geographical objects as purpose, function, composite geographic information constructs, simple geographic information constructs, similarity, observables, and the existence of geographical objects are considered semantic levels. Each semantic level contains one or many domain attributes.

According to identity criteria (Guarino, 1999), hierarchical relations in ontologies have different strength. Relation strength defines the type or role of the geographical objects. Different roles of the same geographical object show the strength of roles by geographic information constructs. Geographic information constructs are more flexible to explore the semantic information than other geographical representations. Considering semantic levels of the proposed framework in (Couclelis, 2010), function

* Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-4-W18-29-2019 | © Authors 2019. CC BY 4.0 License.

level has more conceptual similarity to the roles of geographical objects in formal ontologies.

Significance of semantic levels is a measure of informativeness of tags to explain the POIs. Significance of semantic levels can be calculated by rough set theory as a mathematical tool of indiscernibility and incomplete information systems(Pawlak, 1999). The aim of this paper is to propose an approach to measure the significance of semantic levels to improve the semantic quality of tagsin OpenStreetMap. Finding the most significant semantic level for categories of geographical objects can be applied as the basic information for recommendation tools of the contribution process.

1.2 Semantic levels of geographical objects

According to the semantic levels(Couclelis, 2010), properties for each semantic level describe the way of representation of geographical objects and human configured entities. There are connections between higher semantic levels including purpose and function. A geographical object can afford a set of functions to fulfil the purposes which are defined for. Discovering the semantic levels of geographical objects such as purpose and functionality needs to define clearly the semantic context of geographical objects. Guarino (1999) described different kinds of “is-a” relation which can also apply to geographical objects. Identifying the roles of geographical objects is used to describe the functions. Roles(Guarino, 1999) are non-essential properties that can be confused with type of the geographical objects in tagging process of POIs by contributors.

Purpose as the highest semantic level reveals the contributor’s intentionality. Functions of a geographical object can be explained by the roles which they play for a specific purpose. Composite geographical object is defined to discover the associations between the objects. For example, a campus as a composite geographical object consists of a number of rooms and yard. Simple geographical objects describe the categorization of objects from different points of view such as geometric categorization. Addressing lower semantic levels is beyond of discussion.

Table 1 shows an example of semantic levels of tags about the building which is tagged as a school or a disaster recovery center.Contributors of OpenStreetMap describe POIs in the form of tags which can be corresponding to one, or more semantic levels. An important part of the tags is to describe the POIs as the human configured entities. For instance, the semantic level of geographic information construct for school is more significant in comparison to “polygon” or “geographic region”.

Geographical object: tag “building = school”		
Purpose	School	Disaster recovery center
Function	Education	Disaster response
Composite Objects	Rooms, yard	Rooms, structure
Simple Objects	Polygon	Polygon
Similarities	Building = school	Disaster.shelter_type= logistic

Table 1. Semantic levels for the different contribution of school

1.3 Rough set theory

Indiscernibility between entities is the main result of incomplete information. Describing different geographical objects by the same tags or mixed semantic levels causes indiscernibility and semantic heterogeneity. The significance of semantic levels defines the most important attributesofgeographical objects.Rough set theory is a mathematical approach to extract the clusters of indiscernible objects instead of single and crisp objects (Pawlak, 1999). Indiscernibility between OpenStreetMap categories of POIs is due to using similar tags for different types of POIs or different semantic levels for similar types of POIs. In this research the significance of semantic levels is measured by rough set theory.

Attributes significance is one of the fundamental methods in Rough Set Theory to discover the meaningful attributes of objects in the incomplete information systems. The minimal subset of significant attributes for each type of geographical objects provides the basic semantic context. Attribute significance shows the weight of the attributes and helps to construct the rough ontologies of incomplete information.

Definition 1: Information system *IS (U, A)*is defined as the representation of input databy *U* as the universe of discourse and *A* as the attributes.*U* is a non- empty and finite set of objects and *A*is a non-empty finite set of attributes.

Definition 2:TheEquivalence relation $R \subseteq X \times X$is a binary relation which is :

- reflexive: (xRx for any object x) ,
- symmetric (if xRy then yRx), and
- transitive (if xRy and yRz then xRz).

Definition 3:The equivalence class$[X]_R$ of an element $x \in X$consists of all objects$y \in X$ such that *xRy.*

Definition 4: For *IS (U, A), IND (B)*for $B \subseteq A$is defined as the indiscernibility relation by:

$$IND(B) = \{ \quad (x, y) \in U \times U: F(x, a) = F(y, a), \forall a \in B \quad \} \qquad (1)$$

F is the function which defines the values of attributes for objects *x, y*$\in U$. If the values of an attribute such as *B* for two different classes *x, y*$\in U$are equal, then *x* and *y* will be indiscernible by *B*. The relation is called B-indiscernibility and denoted by *U/ IND (B).*

Definition 5:For*IS (U, A, if* $B \subseteq A$ and $X \subseteq U$ then the *B*-lower and *B*-upper approximation of *X*are defined byB_X and $B^{-}X$ where :

$$B\_X = \cup \{Y_i \in U/IND(B) | \, Y_i \subseteq X \quad \}(2) \qquad (2)$$

$$B^{-}X = \cup \{Y_i \in U/IND(B) | \, Y_i \cap X \neq \emptyset \quad \} \qquad (3)$$

Definition 6:Boundary region of an attribute $B \subseteq A$ is defined as the $B^{-}X$ -B_X and consist of those objects that cannot be classified based on the *B*. If the boundary region of Bis empty then the set will be crisp otherwise it will be rough.

Definition 7:The classification quality $r_B$is defined based on the lower approximation of attributes $B \subseteq A$ by:

https://doi.org/10.5194/isprs-archives-XLII-4-W18-29-2019 | © Authors 2019. CC BY 4.0 License.

$$r_B(F) = \sum_{i=1}^{n} \frac{|B\_(X_i)|}{|U|} \quad (4)$$

2. SIGNIFICANT OF SEMANTIC LEVELS FOR TAGS OF OPENSTREETMAP

In order to study the semantic contents of tags in OpenStreetMap, the ontological framework proposed by Couclelis(2010) is implemented as the reference for information table. significance of semantic levels for tags of buildings in OpenStreetMap are measured based on the principles of rough ontology (Chen & Lv, 2010).

2.1 Information table of tags

Analysis of the semantic problems of tags by rough set theory needs to build the information table of tags. Information table consists of two disjoint sets of attributes as condition and decision attributes.Description of geographical objects in the form of tags is corresponding to one, two or more semantic levels.Tags are assumed to reflect the purpose of the contributors or functions afforded by POIs or etc. Weconsider each semantic level as a condition attribute. So tags could not provide complete semantic details for POIs.

We define the "semantic quality" as the decision attribute to categorize the geographical objects. Decision attributes are applied to define the indiscernibility of objects based on the condition attributes. In Table 2, we present the information table which is composed of the top values of tags[1] (Key = Building) as and semantic levels. If values of tags satisfy the semantic level then the value of the cell in the information table will be defined 1 and otherwise 0.

Value of Building	Decision Attribute	Condition Attributes		
	Semantic quality	Purpose / Function	Composite Geographical objects	Simple Geographical objects
Yes	0	0	0	1
Apartment	0	0	0	1
House	1	1	0	1
Industrial	0	1	0	0
Residential	0	1	0	0
Garage	1	1	1	1
Hut	0	0	1	1
Detached	0	0	1	0
Shed	1	1	1	1
Roof	0	1	0	1
Commercial	0	1	0	0
Terrace	1	1	1	1
School	0	1	1	0

Table 2. Information table of tags for buildings in OpenStreetMap by semantic levels

According to values of semantic quality as the decision attribute, buildings which are tagged as the "apartment" or "industrial" are vague and purpose, function, and other semantic levels need to be described in more detail. The value of "yes" is also a vague value which is very common in OpenStreetMap. This value does not provide any information about categories of buildings. Such value is recognized as the source of heterogeneity in ontological refinement of tags in OpenStreetMap.

2.2 Significance in rough ontology

To calculate the significance of semantic levels of tags in OpenStreetMap, rough model (*U*, *C*, *D*, *F*) of tags of buildings are defined by the set of condition attributes *C*, decision attribute *D*, and *F*. According to the rough set theory, the significance of a condition attribute $C_i \in C$ ($C \subseteq A$) is defined by the classification quality. Significance of an attribute indicates the importance of an attribute in categorization of objects.

$$\text{Significance of } C_{i=}\ r_{C}\ (F) - r_{C\text{-}\{Ci\}}\ (F) \quad (5)$$

Where r_C (F)= classification quality of C
$r_{C\text{-}\{Ci\}}$ (F) = classification quality of C-{C_i}
F = indiscernibility function of C
C = set of condition attributes
C_i= a condition attribute

U as the universe of discourse is including the most used values of tags which are applied to describe the buildings in OpenStreetMap including "yes", "apartment", "house", "industrial", "residential", "garage", "hut", "detached", "shed", "roof", "commercial", "terrace", and "school".

The well-matched semantic levels for buildings are purpose or function, composite geographical constructs, and simple composite geographical constructs. The semantic quality as the decision attribute shows how the tags describe the buildings. The condition attributes helps to show the vagueness of tags for different types of buildings.

By values of semantic quality which are defined as 0 and 1, there are two categories of objects as X_1 and X_2 and classification quality is calculated by (6).

$$r_C(F) = \sum_{i=1}^{2} \frac{|C\_(X_i)|}{|U|} = \frac{(|C\_(X_1)|+|C\_(X_2)|)}{|U|} \quad (6)$$

Where $r_C\,(F)$ = classification quality
$|C\_\ (Xi)|$ = number of categories based on the *F*
X_i = subsets of *U* that $U = \cup_{i=1}^{n} X_i$

$$C\_X = \cup\{Y_i \in U/IND(semantic\ quality)|Y_i \subseteq X\ \ \} \quad (7)$$

Where $U/IND(semantic\ quality)$are the categories of *U* by attribute of semantic quality.

3. RESULTS

The significance of the condition attributes as the semantic levelsprovides a measure that can be applied in recommendation tools to improve the semantics of POIs during the contributionprocess. In this case study, condition attributes are *C= {Purpose/Function (C₁), Composite Geographical Objects (C₂), Simple Geographical Objects (C₃)}*. The significance of three conditions attributes *C₁*, *C₂*, and *C₃* as the semantic levels are calculated. Details of rough set computation for this research are presented in Appendix.

[1] The most common tags of buidingsin TagInfo 2018-07-08

$$\textit{significance of } C_1 = r_C(F) - r_{C-\{C_1\}}(F) = 0.846 - 0.384 = 0.462 \quad (8)$$

$$\textit{significance of } C_2 = r_C(F) - r_{C-\{C_2\}}(F) = 0.846 - 0.538 = 0.308 \quad (9)$$

$$\textit{significance of } C_3 = r_C(F) - r_{C-\{C_3\}}(F) = 0.846 - 0.308 = 0.538 \quad (10)$$

According to the significance values, the hierarchy of semantic levels based on tags and significance of semantic levels is constructed in Figure 1.

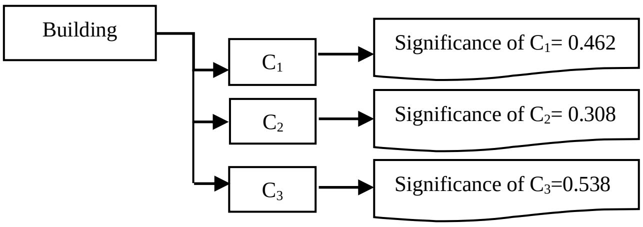

Figure 1.Hierarchy of semantic levels based on the significance of C_1, C_2, and C_3

The maximum significance of most common tags about buildings belongs to the simple geographic information constructsand purpose/function. Describing buildings as composite geographic information construct has lower significance.

4. CONCLUSIONS

Each type of geographical objects can be described in one or more levels based on the ontological framework of semantic levels. Significance of semantic levels shows the power of each semantic level to define the properties of POIs.Finding the significant semantic levels fordifferent types of geographical objects improves the interoperability of VGI with other geographical data sources. Suggesting the significant tags to describe the POIs during the creation or edition of tags will enhance the semantic quality of VGI.

In this paper, the significance of semantic levels is calculated for buildings and results show that simple geographic information construct and purpose/ function are the main semantic levels of buildings.

The significant semantic levels are useful in the creation of the hierarchy of tags for geographical objects in Java OpenStreetMap (JOSM) editor as OpenStreetMappresets. This will result in the improvement of reusability of tags, categorization of geographical objects in OpenStreetMap, and improvement of interoperability.

REFERENCES

Ballatore, A. (2016). Prolegomena for an Ontology of Place. In W. K. Harlan Onsrud (Ed.), *Advancing Geographic Information Science.*

Bishr, M., & Kuhn, W. (2007). Geospatial Information Bottom-Up: A Matter of Trust and Semantics. In W. M. Fabrikant S.I. (Ed.), *The European Information Society. Lecture Notes in Geoinformation and Cartography* (pp. 365-387). Berlin, Heidelberg: Springer.

Chen, H., & Lv, S. (2010). *Study on ontology model based on rough set.* Paper presented at the Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI).

Couclelis, H. (2010). Ontologies of geographic information. *International Journal of Geographical Information Science, 24*(12), 1785-1809. doi: 10.1080/13658816.2010.484392

Frank, A. u. (2001). Tiers of Ontology and Consistency Constraints in Geographical Information Systems. *International Journal of Geographical Information Science, 15*, 667-678.

Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4), 435-459. doi: 10.1111/j.1467-9671.2010.01203.x

Guarino, N. (1999). *The Role of Identity Conditions in Ontology Design.* Paper presented at the COSIT 1999.

Kuhn, W. (2003). Semantic reference systems. *International Journal of Geographical Information Science, 17*(5), 405-409.

Pawlak, Z. (1999). Rough Set Theory And Its Applications To Data Analysis. *Cybernetics and Systems: An International Journal, 29*(7), 661-688. doi: 10.1080/019697298125470

Scheider, S., & Janowicz, K. (2014). Place reference systems: A constructive activity model of reference to places. *Applied Ontology, 9*, 97-127. doi: 10.3233/AO-140134

Scheider, S., Janowicz, K., & Kuhn, W. (2009). *Grounding Geographic Categories in the Meaningful Environment*. Paper presented at the Conference on Spatial Information Theory, Aber Wrac'h, France.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science, 31*(1), 139-167. doi: 10.1080/13658816.2016.1189556

Smith, B., & Mark, D. M. (2001). Geographical categories: an ontological investigation. *15*(7), 591-612. doi: 10.1080/13658810110061199

Vandecasteele, A., & Devillers, R. (2015). Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney & M. Helbich (Eds.), *OpenStreetMap in GIScience,Lecture Notes in Geoinformation and Cartography*. Springer International Publishing Switzerland

APPENDIX

- Categories of values by semantic quality:

F = { {3,6,9,12} , {1,2,4,5,7,8,10,11,13} }

- Indiscernible categories by different condition attributes:

$U / IND\ (C)$ = { {1,2},{3,10},{4,5,11},{6,9,12},{7},{8},{13} }
$U / IND\ (C_1, C_2)$ = { {1,2},{3,4,5,10,11},{6,9,12,13},{7,8} }

$U / IND(C_1, C_3)$ = { {1,2,7},{3,6,9,10,11,12},{4,5,13},{8} }
$U / IND(C_2, C_3)$ = { {1,2,3,10},{4,5,11},{6,7,9,12},{8,13} }

- Classification quality of attributes

$$r_C(F) = \sum_{i=1}^{2} \frac{|C_{-}(X_i)|}{|U|} = \frac{(|C_{-}(X_1)| + |C_{-}(X_2)|)}{|U|}$$

= (|{1,2}|+ |{4,5,11}|+|{6,9,12}|+|{7}|+|{8}|+|{13}|) / |{1,2,3,4,5,6,7,8,9,10,11,12,13}| = 0.846

$r_{C-\{C1\}}(F) =$ (| {4, 5, 11}|+|{ 8, 13}|) / |{1,2,3,4,5,6,7,8,9,10,11,12,13}| = 0.384

$r_{C-\{C2\}}(F) =$ (|{ 1,2,7}|+|{4,5,13}|+|{8}|) / |{1,2,3,4,5,6,7,8,9,10,11,12,13}| = 0.538

$r_{C-\{C3\}}(F) =$ (|{1,2}|+|{7,8}|) / |{1,2,3,4,5,6,7,8,9,10,11,12,13}| = 0.308

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-4-W18-29-2019