Integrating Authoritative and Volunteered Geographic Information - An Ontological Approach

Jimena Martínez Ramos and Rodolphe Devillers

Department of Geography, Memorial University of Newfoundland, NL A1C 5S7
St. John’s, (Canada)

1 Introduction

Many National Mapping Agencies (NMAs) are looking at using Volunteered Geographic Information (VGI), from projects like OpenStreetMap (OSM), to improve their datasets, processes, and reduce their production costs. As a result, integrating geographic datasets is gaining importance [1].

One of the main obstacles to a seamless integration of geographic information is semantic heterogeneity, which refers to the different ways real-world phenomena are conceptualized in datasets [2].

The usual approach for handling semantic heterogeneity between two datasets is by building an ontology for each dataset and then match concepts between the two ontologies. Three main problems arise using this approach. First, ontology matching techniques are not always satisfactory as each dataset has its own data model, so an upper conceptualization is needed to facilitate the matching. Second, this approach requires users to know how to handle ontologies. And third, it does not allow reusing the matching for new datasets, because there is no common ontology.

These issues become even more difficult in the context of VGI datasets, due to their flexible and dynamic data models [3] where users can freely select and add new tags to the map. As a result, semantic heterogeneity may occur inside OSM datasets. One real-world phenomenon can be represented with different tags by different users (synonymy), and users can select the same tag to represent different phenomena (polysysemny).

2 Method

This project aims to develop a method for handling semantic heterogeneity when integrating VGI and authoritative datasets. The method follows two stages.

The first stage consists of building a domain ontology that will serve as common knowledge onto which authoritative and VGI datasets can interface. This should follow standardization guidelines (ISO/TC 211 and OGC), since standards provide the structure for sharing and integrating geographic information [4] and since there is still no standard by which to describe geographic features. We propose using the General Feature Model (GFM) [5] as the starting point to develop the domain ontology. GFM is based on two main features, GF_FeatureType and GF_PropertyType. Both are meta-classes representing, respectively, all feature classes and all their properties. A proof of concept of this approach was done for ‘roads’ and ‘railroads’ features.

The second step is to perform mappings between the datasets and the domain ontology. We choose the R2RML [6] standard, that can be used to express customized mappings from relational databases to RDF [7]. A mapping file based on R2RML is used to specify mappings between each feature class to the concepts of the domain ontology. Such an approach does not require building an ontology for each dataset and the mappings can be reused.

R2RML mapping results in two RDF files populated with the objects that originally were features in the source datasets. After the semantic matching, objects in the output files represent the same phenomena. Once the mappings are defined, the output files are semantically interoperable.

3 Ongoing and Future Work

We developed a proof of concept for the domain ontology and performed the mappings between several datasets but only focused on some feature classes. More work would be needed to create a complete domain ontology following standard guidelines.

For ongoing research, we are qualifying and quantifying semantic heterogeneity in OSM datasets, studying how space and time are affecting semantics in OSM datasets. Such understanding could help improve the quality of semantics in OSM datasets, and in turn help integrating VGI data with other geographic information sources.

4 References