Volunteered and Professional Geographic Information Working in Harmony

Heshan Du[1], Glen Hart[2], Natasha Alechina[1], Mike Jackson[1]

[1] University of Nottingham
[2] Ordnance Survey of Great Britain

The production of Volunteered Geographic Information (VGI) has been seen as a threat to Professional Geographic Information (PGI) such as that produced by national mapping agencies. Though VGI is typically more variable, not as reliable and structured as the PGI, it often reflects changes in the real world more quickly [1]. However, VGI and PGI do not have to be seen as competitive but as complementary, and from a consumer viewpoint the challenge is how to combine these data successfully. This need has been recognised by Ordnance Survey, Great Britain's national mapping, who are working with the University of Nottingham to investigate how such harmony can be obtained through the application of ontologies.

A number of geographic ontologies have been developed independently by different communities for PGI and VGI datasets and the approach we have taken is to utilise the ability of an ontological approach to aid the harmonising of data from these disparate geospatial information sources. We propose a new semi-automatic method to harmonise disparate geospatial ontologies, using description logic reasoning and domain experts' knowledge. The basic principle is, after linking entities (including concepts, roles and instances) from disparate geospatial ontologies, the overall information should be logically coherent and consistent. It is assumed that, an axiom is a premise, which is believed all the time, if it exists in an original ontology, otherwise it is an assumption, which is believed by default, but may be retracted later. Assumptions, in this context, include disjointness axioms, terminology mappings, and instance mappings. Users are allowed to retract or enable existing assumptions, and add new assumptions, during the harmonisation process. A disjointness axiom states that two or more classes are pairwise disjoint, having no common elements. For example, $Person$ and $Place$ are disjoint. The disjointness axioms in ontologies are very useful when detecting logical errors. However, most of the existing geospatial ontologies are underspecified or light-weighted, within which disjointness axioms are not always available. To deal with this, a set of disjointness assumptions will be generated for each input ontology, ensuring that they will not cause any incoherence. To harmonise two ontologies at the terminology level, an initial terminology mapping is calculated using lexical matching. Applying this mapping directly may lead to the incoherence of the overall information. When incoherence is detected by a description logic reasoner, assumptions that cause it, will be visualised clearly to facilitate domain experts taking repair actions, for example removing an assumption. This process iterates until the information is coherent.

Similarly, when harmonising two ontologies at the instance level using an initial instance mapping, if any inconsistency is detected, domain experts are consulted to take repair actions, given assumptions causing it. With respect to special properties of geospatial information, the initial instance mapping will be refined before this consistency checking. Three main constraints are defined, these are, geometry, lexical and cardinality. Firstly, a correspondence is invalid if the geometries of the linked individuals are not matched, for example disjoint. Secondly, a correspondence is invalid if the meaningful labels indicating identity, cannot be lexically matched. The lexical matching here should be robust enough to tolerate partial differences in labelling. For example, a full name and its abbreviation should be matched. Thirdly, if an individual is involved in several different "sameAs"correspondences, then these correspondences need to be repaired, for example, changing the relation from "sameAs"to "partOf". The three constraints complement each other to cope with the following possibilities. Different geospatial individuals may

share the same label or the same location in an ontology. In addition, a same geospatial individual may be represented as a whole in one ontology, whilst as several parts of it in the other.

The Ordnance Survey of Great Britain (OSGB) Buildings and Places ontology [2] and the OpenStreetMap (OSM) controlled vocabularies [3] are employed to evaluate the method. Currently, OSM has not established a standard ontology, but maintains a collection of commonly used tags for main map features [3]. An OSM feature ontology was generated automatically from the existing classification of main features. Both ontologies are written in the OWL 2 Web Ontology Language [4]. The OSGB Buildings and Places ontology has 692 classes and 1230 logical axioms, while there are 663 classes and 677 logical axioms in the OSM ontology. Both ontologies, containing no disjointness axioms, are coherent. Some examples are discussed below, where $\sqsubseteq$ means inclusion , $\equiv$ means equivalence, $\neg$ means complement, and $\in$ means membership.

Example 1. Given,

$$OSM : Roof \sqsubseteq OSM : Building \tag{1}$$

The existence of the following assumptions will lead to incoherence.

$$OSGB : Building \equiv OSM : Building \tag{2}$$

$$OSGB : Building \sqsubseteq \neg OSGB : Roof \tag{3}$$

$$OSGB : Roof \equiv OSM : Roof \tag{4}$$

Domain experts are consulted to decide which assumption(s) to retract.

Example 2.

$$OSGB : 1000002308476718 \in OSGB : HealthCentre \tag{5}$$

$OSGB : 1000002308476718$ refers to a $OSGB : HealthCentre$ labelled as "SNEINTON HEALTH CENTRE".

$$OSM : 62134030 \in OSM : Clinic \tag{6}$$

$OSM : 62134030$ refers to a $OSM : Clinic$ labelled also as "SNEINTON HEALTH CENTRE". Their geometries are very similar. However, the existence of the following assumptions can lead to inconsistency.

$$OSGB : 1000002308476718 \equiv OSM : 62134030 \tag{7}$$

$$OSGB : Clinic \equiv OSM : Clinic \tag{8}$$

$$OSGB : Clinic \sqsubseteq \neg OSGB : HealthCentre \tag{9}$$

Domain experts are consulted to decide which assumption(s) to retract. In this case, to keep the individual correspondence (7), it is reasonable to retract (9).

In this paper, a new semi-automatic logic-based method is proposed to harmonise disparate geospatial ontologies. In future work, we plan to finish the basic implementation of the system, improve its efficiency by employing a truth maintenance system, and design a graphical user interface facilitating users' decision making processes.

References

1. Jackson, M.J., Rahemtulla, H., Morley, J.: The synergistic use of authenticated and crowd-sourced data for emergency response. In: 2nd International Workshop on Validation of Geo-Information Products for Crisis Management (VALgEO), Ispra, Italy (11-13 October 2010)
2. Hart, G., Dolbear, C., Kovacs, K., Guy, A.: Ordnance Survey Ontologies. http://www.ordnancesurvey.co.uk/oswebsite/ontology (2008)
3. OpenStreetMap: The Free Wiki World Map. http://www.openstreetmap.org (2012)
4. W3C: OWL 2 Web Ontology Language. http://www.w3.org/TR/owl2-overview (2009)