What OWL has done for geography and why we don't need it to map read

Catherine Dolbear[1], Glen Hart[1] and John Goodwin[1]

[1] Research & Innovation, Ordnance Survey, Romsey Road, Southampton SO16 4GU, UK
{Catherine.Dolbear, Glen.Hart, John.Goodwin}@ordnancesurvey.co.uk

Abstract. This position paper describes our experiences of authoring a geospatial ontology in the domain of hydrology topography, giving examples of where we find the additional functionality of OWL 1.1 useful. We also comment on approaches to combining spatial and semantic reasoning. We contend that spatial reasoning can be best achieved through a combination of semantic and spatial database technologies working together and spatial relationships expressed in OWL need to represent more than just spatial logics.

Keywords: OWL 1.1, spatial reasoning, geospatial semantic web.

1 Introduction

Ordnance Survey, the national mapping agency of Great Britain, maintains a continuously updated database of the topography of Great Britain, including around 440 million features capturing everything from forests, roads and rivers down to individual houses, garden plots, and even pillar boxes. We are faced with the problem of wishing to exploit the semantics of a very large spatial database. We have developed a subset of the topographic ontology we are building to describe the knowledge we have about the domain and the data we collect. This hydrology ontology currently contains 301 classes and 162 properties, and represents an estimated one tenth of the size of the full topographic ontology. Section 2 uses examples from this hydrological topographic ontology to demonstrate where we would find OWL 1.1 useful and section 3 comments on approaches to combining spatial and semantic reasoning. Practical examples are used to illustrate our contention that spatial queries are better answered by reference to instance data in the database, and spatial relationships at the ontological level, when needed, are often combinations of semantics and spatial logics such as RCC8 [1].

2 OWL 1.1

In this section we examine how useful the constructs introduced by OWL 1.1. [3] are for representing topography. Reflexive, irreflexive and anti-symmetric constructs are certainly of use to our domain. For example the property "flows into" is both irreflexive (any one river cannot flow into itself) and anti-symmetric (if one river flows into the second, the second one can't flow into the first). The spatial relation "adjacent to" is another example of an irreflexive property, although as discussed in section 3, we don't believe there is much to be gained from specifying qualitative spatial relations at the ontology level. Our approach to ontology authoring is for the domain expert to control and author the human readable "conceptual aspect" of the ontology using structured English [2], and then for a knowledge engineer to convert the structured sentences to OWL. However, it is difficult for the domain expert to understand, distinguish between and easily use these constructs .For example, while the "part of" relationship is classically considered to be reflexive, the domain expert would not naturally consider specifying this in an ontology. We have relied on the knowledge engineer to bring them to the domain expert's attention.

Despite it being just "syntactic sugar", *DisjointUnion* is another useful addition to OWL 1.1. We have set it as the default for the domain expert, as most people assume a closed world – meaning that concepts they've specified are automatically different from each other, unless otherwise stated, so it crops up a lot in our ontologies. In some cases, qualified cardinality restrictions can be useful. For example, a confluence connects at least two rivers or streams or canals. However, for many geographic examples, qualified cardinality restrictions or a concrete domain of numbers wouldn't help much, because geography is a domain that contains vague concepts whose distinctions cannot be made explicit. For example, a braided river can only be defined as a river which flows in "several" channels – there isn't a specific minimum number of channels required for it to be called braided. Another example is that we can't draw a distinction between a river and a stream based on an absolute size, since none exists.

We did consider taking advantage of the user-defined data-type functionality of OWL 1.1 to create a spatial data-type to store co-ordinate arrays of the geographical features in our database converted from Oracle SDO_GEOMETRY. Section 3 explains our reasons for rejecting this option. We do have a use for the additional comments functionality of OWL 1.1 however, as we want to link the OWL ontology back to the human-readable conceptual ontology as much as possible. Having comments interspersed anywhere in the OWL ontology means that we could track the conversion from the conceptual ontology structured English sentences into OWL and explain knowledge modelling design decisions. We see this type of documentation as being very important in the long term for ontology content accuracy evaluation and reuse.

A last point about OWL 1.1 – we need tools for authoring the OWL 1.1 ontologies and reasoning over them before we can really demonstrate how useful this extra functionality is to geospatial applications. As OWL grows more complex, with new additions to the standard, we think it's really important that editing tools keep apace – not just for knowledge engineers, but also for domain experts to properly understand the subtle differences of what can be expressed. The new additions to

OWL in 1.1 have understandably mainly been prompted by ease of implementation, but we want to caution that future releases need to be driven first and foremost by what users need, and an appropriate level of abstraction. Since our maps are used for many different purposes, Ordnance Survey is in the difficult position of not knowing the end use of our data or ontologies. The level of abstraction we have settled on is the vernacular – terms and detail that the man in the street will understand, leaving an end user with a generic geographic framework which can be reused and extended for many different purposes. The question we then have of the new OWL 1.1 constructs is – do they make reuse and extension easier or more difficult?

3 Spatial reasoning

Commenting on the future of the semantic web, Tim Berners-Lee recently asserted that "we need to look at existing databases and the data in them" [4]. Central to all Geographic Information Systems (GIS) and more lately to spatially enabled databases systems, is the support of various spatial operators designed to determine the relationship between geospatial objects. For example, all will support (under various names) algorithms to determine spatial containment, overlap, disjointness and so on. Significant work has been done to formalise these relationships such as the 9 intersection model [5] and RCC8 [1]. Due to their formal expression, these calculi can be expressed as topological relationships using OWL and thus enable reasoners to perform inference over spatial data based on topology. As an example, RCC8 has been implemented using OWL to support an experimental ontology editor [6].

However, a very significant proportion of all geospatial data does not contain explicit semantics describing topological information. It is more usual for these systems to determine specific topological relationships through geometrical calculation on data with positional information. For example it is unlikely for a geospatial database to explicitly contain the topological relationship that a specific house is "contained within" a specific garden. Rather, the SQL (Structured Query Language) spatial containment operator will be used to test whether the building instance is indeed contained within the garden. Such approaches are necessary because it would be impractical to explicitly pre-compute all topological relationships between all the objects in a geospatial database. This means that existing reasoners are unable to perform topological inference on the majority of geospatial data as they are unable to compute the necessary topology. Thus the DL reasoning is limited to qualitative reasoning over the ontology rather than quantitative reasoning over instance data. As we are trying to serve the GIS community, with its much larger and more complex spatial databases, our work differs from those approaches which use RDF triple stores to store geospatial data. Here, relatively small RDF stores are used, where the "geospatial" element is limited to location, and some spatial relations that have been captured via annotation, web scraping or pre-computation using a subset of the SQL spatial operators. Qualitative spatial reasoning at the ontology level might be useful in this case; however, this isn't what we're dealing with.

RCC8 and SQL spatial operators define very specific relationships between geometries, such as (using the SQL terminology) touching, contains, inside etc.

However, these rarely correspond to the *semantic* spatial relationships that domain experts would naturally use, or worse, just imply. For example, we might specify that a “field is surrounded by a wall”. Assuming, for the sake of argument, that the wall is represented as a polygon not a line, we might map the “surrounded by” property to the SQL INSIDE operator (non-tangential proper part, in RCC8 speak). However, in most cases there could easily be a break in the wall, implying SQL COVERS (tangential proper part) – but it wouldn’t really affect our semantic understanding, that is, how we thought about the field being surrounded by the wall. A relationship like “next to” may be equally problematic. Although both the 9 intersection model and RCC8 have well defined and predictable “next to” relationships (SQL TOUCH or RCC8 externally connected), a “badly behaved brother” also exists. Given a situation where a house is surrounded by a garden which in turn is physically next to a foot path running by the side of a road, if asked “Is the house next to the road?” most people would answer “Yes.” A GIS or RCC8 reasoner would say no, since the house is not physically located next to the road. Thus a version of “next to” exists where geospatial objects that are deemed to be insignificant are filtered out. So it would be useful to use the ontology to specify which spatial operators we really meant. This would be particularly useful for even vaguer spatial relations like “near to”.

4 Conclusions

While we welcome the additional functionality of OWL 1.1, and have described some situations where we can use it in a hydrology ontology, it’s not yet clear to us how much additional reasoning power it gives us over OWL 1.0, until tools are available. We’d caution against OWL becoming too “bloated” as, for example with spatial reasoning, it may be better to research into the interface between OWL and other representation formalisms (such as rules or links into SQL spatial operators) rather than extending OWL itself to encompass spatial reasoning.

References

[1] D.A. Randell, Z. Cui and A. G. Cohn, A spatial logic based on regions and connections. In Proc. 3rd International Conference on Knowledge Representation and Reasoning pp165-176 San Mateo, (1992), Morgan Kaufman

[2] A Methodology for Building Conceptual Domain Ontologies K. Kovacs, C. Dolbear, G. Hart, J. Goodwin and H. Mizen Ordnance Survey Research Technical Report I-002 2006.

[3] OWL 1.1 Web Ontology Language Overview P. Patel-Schneider, http://owl1-1.cs.manchester.ac.uk/Overview.html

[4] B. Runciman, Interview with Tim Berners-Lee, ITNOW, British Computer Society, 2006

[5] Shariff A, R Egenhofer M J and Mark D M, Natural Language spatial relations between linear and areal objects: The topology and metric of English Language terms.(1998) Int Journal of Geographical Information Science 12(3): 215-246

[6] Smart, P, Abdelmoty and Jones, C. A Visual Editor for Validating Geo-ontologies in OWL, Conference Proceedings GISRUK 2006