Using ontologies with case studies: an end-user perspective on OWL

J. Gary Polhill, Macaulay Institute, Aberdeen, AB14 0PR. g.polhill@macaulay.ac.uk

Gina Ziervogel, Department of Environmental and Geographical Science,
University of Cape Town, Private Bag x3, Rondebosch 7701, South Africa. gina@egs.uct.ac.za

This document describes work undertaken to develop an OWL ontology of a case study of farmers in South Africa. This work was done with a view to seeing whether or not such an ontology would be of use in working with qualitative and/or quantitative case study data. For example, could the ontology be used to check whether the data were being interpreted consistently? We describe the process by which the ontology was developed and offer some reflections on tools and utilities that would make this easier and the strengths and weaknesses of OWL for representing concepts in a real-world sociological case study.

1 Scenario and source material

In 2002/03 there was a food-security crisis in South Africa. One of the issues identified was the availability of long-range climate forecasts to subsistence farmers. Farmers can use these climate forecasts to adopt strategies that mitigate against climate stress, such as planting drought-resistant cultivars. However, climate forecasts are aimed largely at the commercial farming sector. Subsistence farmers, who tend to be women supporting their families while their husbands work in the city, are less likely to be part of the climate information dissemination network.

In developing an OWL (Antoniou & van Harmelen, 2004) ontology of this case study, we drew mostly on three papers. One, by Emma Archer (2003), focused on issues in climate information dissemination worldwide, whilst the second (draft) paper (Archer & Easterling, n.d.) discussed the South African climate information dissemination system in particular, identifying areas in which it could be improved. The third paper (also a draft) (Ziervogel & Bharwani, n.d.) compared various strategies employed by farmers in a village in the Limpopo province of South Africa to adapt to multiple stressors (including climate) influencing their livelihoods.

2 Ontology development methodology

The ontology was prepared through considering its possible uses. Since OWL allows representation of individuals, the ontology could be used to store case study data from interviews and questionnaires. Reasoning services could then be applied to the source data according to qualitatively defined concepts (such as ‘vulnerable farmer’), allowing automatic consistent classification of individuals. The source material was therefore used to look for and, where possible, define such concepts as an initial stage in contributing to such an analysis.

Automatic ontology learning from source texts is an on-going research area, which Gómez-Pérez & Manzano-Macho (2005) conclude has yet to establish a detailed methodology (p. 207), and none of the tools available are able to evaluate their accuracy (p. 208). That said, the methodology used here was a seven-stage manual process outlined below, which could no doubt have been facilitated by existing tools. The ontology itself was built using Protégé[1].

1. *Assembling source material*.

[1] http://protege.stanford.edu/

2. *Deciding high-level classes*. The source material was searched for the key concepts and central themes that would suggest the classes and relationships that would be needed in the ontology. In this case, these were such things as farmers, adaptation, and resilience.
3. *Detailed analysis*. Again, using the source material, instances of the high-level classes were highlighted, and linguistic modifiers (e.g. adjectives) of them analysed to suggest subclasses, and datatype or object properties. Several subclasses of farmer were identified by this process, including `WomenFarmers`, and `SmallholderFarmers`. From the verbs in the sentences using these terms, it was possible to find out differences among the various classes and areas of commonality, e.g. that farmers choose the crops they will plant. One issue that became apparent in this stage of the process was the difference between ontological aspects of the scenario and research findings. For example, in gender analysis it might be that one starts with the assumption that women are disadvantaged (i.e. this fact is part of the ontology). Equally, however, one might prefer to define what it means to be disadvantaged and subsequently find that members of this class are predominantly women. Confusing the two would be undesirable as it could lead to a tautology.
4. *Developing support classes*. Support classes, for the purposes of this paper, are classes that are not defined as part of the scenario, but are assumed as common knowledge. Examples are social networks, and concepts of geographical area, in particular remoteness, which is critical to an understanding of vulnerability in this case study. Ideally, such classes would be imported from other ontologies, something OWL is specifically designed to do. However, in practice, searching for ontologies to import (e.g. using engines such as Swoogle[2]) proved too laborious: the tools are not yet mature enough to enable users to quickly sift through the results to determine which are appropriate.
5. *Class definition*. The distinction between primitive and defined classes in OWL is significant when it comes to the application of reasoning services, since only defined classes can be directly inferred to have members.
6. *Testing*. The ontology can be tested using some built-in features of Protégé, as well as checking for concept consistency, and checking the classification of test individuals using the reasoner (RACER). Data from the questionnaires can also be entered into the ontology, to check for the absence of concepts that applied to the gathering of evidence, but did not appear in the case study texts.
7. *Refinement*. The ontology is refined by iterating the process to bring in extra source material, respond to the results of the tests, or import existing ontologies for support classes.

3 OWL: an end-user perspective

Ontology development is quite a time-consuming process. Several attempts were required to get to a working ontology. We identified the following issues with the tools used to create the OWL:

- Protégé makes it easy to create an ontology that becomes OWL-Full (e.g. through creating transitive functional properties), and although a facility is provided to allow the user to check the species of OWL used in the ontology, it does not specify *why* the language subtype is as found to allow the user to make adjustments accordingly.
- Similarly, a lack of explanation from the reasoner as to *why* a particular ontology has been found to be inconsistent, or *why* an individual has been classified as belonging to a particular concept, makes it difficult to track faults in the ontology, and in some cases, led to abandonment of the current version when perhaps this need not have been necessary.
- The ontology does not stand alone as a document, and needs a user-guide to provide an overview and explain the various definitions. Perhaps there could be some metadata to

[2] http://swoogle.umbc.edu/

help users of an ontology get to grips with it (e.g. 'start here'), or to facilitate the automatic creation of annotated navigations through an ontology, which is the kind of material a user-guide might contain. It would certainly be useful to have some standard metadata (akin to Dublin Core[3]) to assist with providing links between assertions in the ontology and source materials.

For OWL itself, there is the question of whether it is expressive enough whilst still allowing decidable reasoning. There are various concepts we found difficult to describe in OWL. The Archer & Easterling paper, for example, refers to centralised top-down social networks. In idealised form, such networks consist of a single root-node and directed links that form a strict tree-like structure to the leaf nodes. The following are necessary and sufficient conditions for such a network:

(a) All nodes except the root node have exactly one input link.
(b) All nodes except leaf nodes have one or more output links.
(c) A root node has no input links.
(d) There is exactly one root node in the network.
(e) All links are between different nodes.

Conditions (a), (b) and (c) are relatively easy to define using cardinality restrictions in OWL-DL. Classes of node for leaf nodes and the root node can also be defined. Condition (d) is more challenging, as there does not seem to be any way of saying in OWL that only one member of the nodes of a network may be a root node. In the ontology we worked round this by introducing to the `Network` class an additional subproperty of `hasNode` called `hasRootNode`, and using a cardinality restriction on that. However, this would still allow a network containing root nodes in the range of `hasNode` only one of which is asserted to be in the range of `hasRootNode` to be wrongly recognised as an instance of a centralised top-down network. Condition (e) is more difficult still, as there is no way (that we could see) to assert that the individuals in the ranges of two properties (`hasFromNode` and `hasToNode` in this case) must be different.

4 Future work

Future work will continue to look at the ways in which ontology reasoning services can augment existing research practices in the social sciences. Can ontologies be used to facilitate questionnaire design or data entry, as well as analysis? Can ontologies be used to bridge the gap between case study evidence and representation in social simulations? What tools do social scientists need to facilitate the useful application of ontologies in their work? What languages are appropriate for representing ontologies in the social sciences?

5 Acknowledgements

This work was funded by the Scottish Executive Environment and Rural Affairs Department, and by the EU Framework Programme 6 NEST Pathfinder Initiative on Tackling Complexity in Science.

6 References

Antoniou, G., & van Harmelen, F. (2004) Web Ontology Language: OWL. In Staab, S., & Studer, R. (eds.) *Handbook on Ontologies*. Berlin: Springer, pp. 67-92.

[3] http://dublincore.org/

Archer, E. R. M. (2003) Identifying underserved end-user groups in the provision of climate information. *Bulletin of the American Meteorological Society* November 2003, pp. 1525-1532.

Archer & Easterling (n.d.) Mitigating climate stress: Strengthening the ‘end-to-end’ climate information system in South Africa. *Draft paper to be submitted to Environment*.

Gómez-Pérez, A. & Manzano-Macho, D. (2005) An overview of methods and tools for ontology learning from texts. *The Knowledge Engineering Review* 19 (3), pp. 187-212.

Ziervogel, G., & Bharwani, S. (n.d.) Adapting to variability: Pumpkins, pumps, poverty and the role of climate. *Draft paper dated February 2004*.