OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies

Conference or Workshop Item

OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies

Allan Third and Sandra Williams and Richard Power

Computing Department, Open University, UK
{a.third,s.h.williams,r.power}@open.ac.uk

Abstract. It has frequently been observed that domain experts are not necessarily ontology experts, and that the production of ontologies would be aided if they could read and edit axioms in natural language. The *SWAT Tools Verbaliser* is available, via a web interface, for verbalising OWL ontologies as texts in a controlled fragment of English. Taking as input any OWL ontology, the verbaliser creates a lexicon containing entries for all the entities in the input, and uses it to generate an English sentence corresponding to each logical axiom. These sentences are then organised into a document structure similar to that of an encyclopaedia, with an entry providing a definition, typology and examples for each entity. The output is either organised, easily-navigable English text encoded in XML, or a copy of the input OWL in which each entity is annotated with its description entry. The generated texts have been evaluated in a number of ways which are briefly presented here.

1 Introduction

There are many reasons for wanting to translate the formal axioms of an ontology into a natural language such as English. First and foremost is the oft-observed fact that domain experts typically do not know ontology languages such as OWL[10], nor have the time or inclination to learn them. Even where knowledge of OWL and of a particular domain coincide in the same person, the task of checking the correctness and coverage of an ontology ought to be greatly assisted if its contents are at least readable by a second person (or more) with knowledge of the domain in question. Once authored and satisfactorily verified, a natural-language "verbalisation" of an ontology can provide a useful "way in" for human readers seeking to find out about its contents.

Some large-scale ontology building projects, such as the OBO Foundry [6], even have a requirement that every axiom have an NL representation. As these can be very large (the Gene Ontology [8], for example, contains 34812 terms at the time of writing), and are generated automatically, it is clearly impractical to write these NL representations by hand. An automatic solution would clearly be beneficial. As the primary aim of these texts is to be read by humans, it is desirable that they be generated and presented in a form which suits human comprehension, navigation and readability needs. The task of translating ontology axioms to a natural language such as English ("verbalisation") has been pursued by several research groups [1, 2, 5].

The SWAT project is concerned with the development of tools for editing OWL ontologies in natural language. One component of our research therefore focuses on the task of natural language generation from OWL, so that existing ontologies can be imported into an NL editing environment, and, indeed, so that "roundtrip" editing can take place – from English to OWL back to English again – in successive iterations of the editing process. We present here the verbaliser tool we have developed, available on the web at [9]. The verbaliser is written in Prolog, and has a multi-stage pipeline architecture, in which the output(s) from previous stages can serve as the input(s) to later ones. Key stages are: lexicon construction, document structuring and verbalisation. Output is not only organised, easily-navigable English text encoded in XML, but also OWL, annotated with generated English class descriptions. Compared with other verbalisers, our tool offers: more sophisticated rules for deriving lexicon; better text organisation; and output crosslinked to OWL. A detailed comparison can be found in [7].

2 From OWL to Text encoded with XML

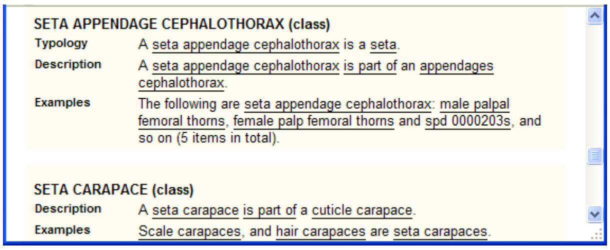

Fig. 1. XML output viewed in a web browser as structured text.

Lexicon construction OWL entities may have machine-readable identifiers (IRIs) or human-readable labels. To generate readable text, it is necessary to construct a *lexicon* of (in this case, English) words or phrases such as "pizza topping", linked to the IRI from which it was generated. We do this by applying heuristics to infer an English word or phrase from labels or abbreviated IRIs. We relate OWL entities to grammatical categories in a straightforward way: named individuals become proper nouns, class names, noun phrases, and so on.

Document structuring It is important to structure the output text in a manner that aids navigation. Our output text is structured at a number of levels as shown in Figure 1. The first (or top) level organises ontology entities (classes, properties and individuals) alphabetically as headings. A second level arranges relevant axioms under sub-headings (Definition, Taxonomy, Description, Distinctions and Examples) according to their logical type [12]. A third level of organisation aggregates axioms with identical structures and one identical argument in a single sentence [11]. This can lead to very long lists of subclasses or members of a class, so these are truncated to a predefined maximum length.

Verbalisation of axioms We take the position (as is common in ontology verbalisation) that each OWL functor corresponds to a particular kind of sentence. For example, a subclass statement `SubClassOf(`A`,` B`)` corresponds to a sentence of the form "An/Every A is a B". The verbaliser contains a grammar which, using the lexicon, generates sentences from axioms according to these correspondences. Figure 1 gives some examples of these verbalisations.

3 Output

As the structured output contains a paragraph of text for each entity in the ontology, it is possible to embed the verbaliser output into the original ontology file as annotations on entities. This produces a file which can be loaded into standard OWL tools (such as Protégé [4]) and manipulated as normal. Figure 2 shows an example text displayed in Protégé.

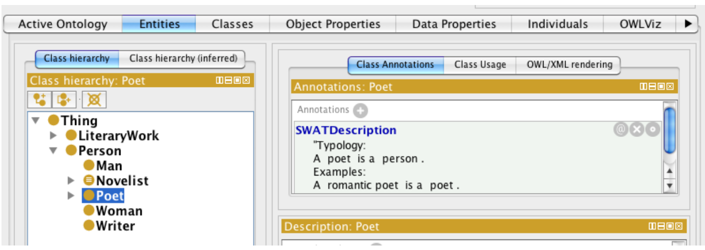

Fig. 2. Verbaliser output annotating an entity in Protégé

4 Evaluations

We have evaluated the verbaliser output using surveys to assess the fluency of the generated text and how well it conveys the intended meanings. These were conducted with developers of the Experimental Factor Ontology (EFO) [3]. The texts were found to convey the intended meanings satisfactorily. The surveys also revealed that more "natural" English was preferred over excessively formal or "OWL-like" English, even where such formal language made the underlying OWL semantics more explicit. One encouraging consequence of these surveys was that an error in the EFO was identified via our generated texts, and corrected. A full description and discussion of these surveys can be found in [7].

Text organisation was evaluated through a navigation task in which two groups of participants were asked to locate information in texts generated by our system and judge how difficult it was to find. One group searched for information in a structured text (see figure 1) the other in an unstructured text (where axioms in the ontology were verbalised as single sentences presented in a long list). The two texts were generated from the same ontology thus they contained the same information. Results supported our assumption that people using the structured text would judge the tasks easier, although performance was similar on both texts. For full details, see [12].

5 Availability

The verbaliser is available online at [9]. It is possible to select different outputs, including annotated OWL, or an XML-based hypertext. An advantage of the XML format is that it preserves explicit links between the English text and the original OWL axioms and entities, making it useful in future for interfaces with, for example, side-by-side English and OWL editing, allowing, among other things, a user to copy/paste English and have the underlying OWL copied in the background. The verbaliser will be demonstrated with a sample of logically-complex ontologies, showing the range of outputs and their functionality (e.g., hypertext structure, Protégé compatibility).

References

1. Cregan, A., Schwitter, R., Meyer, T.: Sydney owl syntax - towards a controlled natural language syntax for owl 1.1. In: OWLED (2007)
2. Kaljurand, K., Fuchs, N.E.: Verbalizing owl in attempto controlled english. In: Proceedings of Third International Workshop on OWL: Experiences and Directions, Innsbruck, Austria (6th–7th June 2007). vol. 258 (2007)
3. Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A., Parkinson, H.: Modeling sample variables with an experimental factor ontology. Bioinformatics 26(8), 1112 (2010)
4. Protégé: (August 2011), `http://protege.stanford.edu/`
5. Smart, P.R.: Controlled natural languages and the semantic web (July 2008), `http://eprints.ecs.soton.ac.uk/15735/`
6. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L., Eilbeck, K., Ireland, A., Mungall, C., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 25(11), 1251–1255 (2007)
7. Stevens, R., Malone, J., Williams, S., Power, R., Third, A.: Automating generation of textual class definitions from owl to english. Journal of Biomedical Semantics 2(Suppl 2), S5 (2011)
8. The Gene Ontology: (August 2011), `http://www.geneontology.org/GO.downloads.ontology.shtml`
9. The SWAT Project: (August 2011), `http://swat.open.ac.uk/swat`
10. W3C Consortium: (April 2010), `http://www.w3.org/TR/owl2-syntax/`
11. Williams, S., Power, R.: Grouping axioms for more coherent ontology descriptions. In: 6th International Natural Language Generation Conference. pp. 197–202 (2010)
12. Williams, S., Third, A., Power, R.: Levels of organisation in ontology verbalisation. In: Proceedings of the 13th European Workshop on Natural Language Generation (forthcoming) (2011)