Improving Ontology Verbalization using Semantic-level Refinement (Extended Abstract)

Vinu E. Venugopal[1] and P. Sreenivasa Kumar[2]

[1] Faculty of Science, Technology & Communication
University of Luxembourg, Luxembourg
vinu.venugopal@uni.lu,
[2] Department of Computer Science and Engineering,
Indian Institute of Technology, Madras, India
psk@cse.iitm.ac.in

Verbalization tools such as ACE [1], NaturalOWL [2] and SWAT Tools [3], are generally used for generating (controlled) natural language (NL) descriptions of logical statements from a given W3C Web Ontology Language (OWL) based ontology, making the underlying knowledge accessible to non-ontology communities as well. In our research, restricting our scope to NL definitions of *individuals* and *atomic concepts*, we found that the approaches currently followed in the existing tools typically consider all logical conditions that are associated with the entity (an individual or a concept) and translate them into a bunch of NL descriptions. But the verbatim fidelity of such descriptions to the underlying OWL statements makes them a poor choice for expressing the underlying knowledge. Especially, for a domain expert who is having limited knowledge about the formal representations of ontology, the generated descriptions would be always redundant and misleading. In an ontology authoring context that involves collecting feedback from these experts, any incorrect feedback would have a negative impact on the construction of the domain ontology. This issue had been previously observed and reported in papers such as [4, 3, 5, 6], where the authors have tried to make the generated definitions closer to the actual definitions of the entities by applying operations such as grouping and aggregation over the verbalized texts. But, since this refinement is happening only at the NL text level, the opportunity for a logical-level refinement of the OWL statements to generate a more concise and human-understandable representation has been ignored completely.

A notion for removing redundancies in a verbalization context was first introduced in [7], where the authors have clearly established the fact that omitting "obvious" axioms while verbalization leads to a better reading experience for a human. By "obvious" axioms the author means those axioms whose semantics are in some sense obvious for an average human reader. (For example, phrases such as "junior school" explicitly convey the meaning that a junior school is a school.) In our work, we go further and establish that more inference-based redundancy removal could be performed (similar to the notions proposed in [8]) than removing just the morphological variants of the entity names, for greatly improving the quality and intelligibility of verbalized text.

In what follows, we demonstrate a brief example showing the flavor of our proposed approach. Let the three axioms (from People & Pets ontology [9]) given

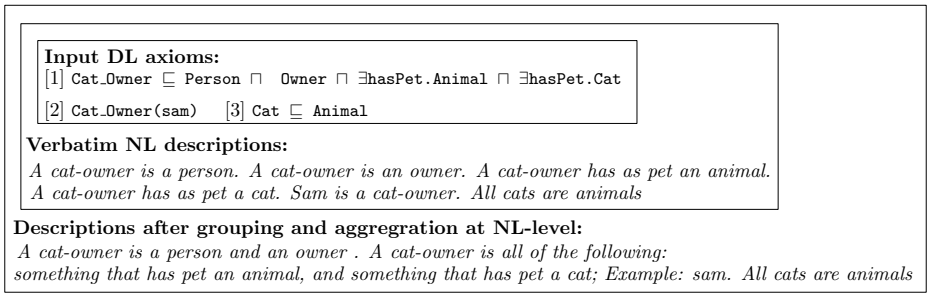

Fig. 1. An example showing two variants of verbalized texts

in Figure 1 be the axioms to be verbalized. The two variants of the NL descriptions defining the individual `sam` are shown in the outer boxes of the figure. We could easily observe that these descriptions contain redundant information even after grouping and aggregation.

As mentioned above there are different types of redundancies one can observe in a verbalized text. The obvious type is the repetition of linguistically similar texts; for example, a `Cat_Owner` is an `Owner`. Refinements of this redundancy is often performed optionally during verbalization. This is because it could adversely affect the correctness of the definition. Another type of redundancy includes those generic restrictions which can be logically inferred from more specific restrictions; for example, having said "A cat-owner has as pet a cat.", it is not necessary to say "A cat-owner has as pet an animal." This paper deals with removing complex redundancies of the latter kind using a set of carefully framed $\mathcal{SHIQ}$ description logic based rules which are repeatedly applied on the selected restrictions to attain a refined form of restrictions. More details about the refinement rules and the algorithms involved in the application of the rules can be found in [10], a technical report.

For the illustrative input axioms, our approach could produce a description similar to: *Sam: is a cat-owner having at least one cat as pet*; such that the redundant portion of the text *has as pet an animal* (since it clearly follows from *having at least one cat as pet*) is removed.

In the technical report that we have mentioned above, we provide the proofs for the correctness of all the refinement rules, which guarantees that the semantics of the logical statements before and after the refinement would be the same. In addition, we did a detailed empirical study to validate the following two propositions by using case studies and by employing statistical significance tests. Firstly, logical-level refinement could significantly improve the intelligibility of the domain knowledge when expressed in NL. Secondly, such NL definitions can be effectively used for validating the correctness of ontologies. It is also observed that our approach is especially beneficial for domain-experts in improving their intelligibility of the domain from the generated texts on comparing with those generated using the existing tools.

References

1. Kaljurand, K., Fuchs, N.E.: Verbalizing OWL in Attempto Controlled English. In: OWLED. Volume 258. (2007)
2. Androutsopoulos, I., Lampouras, G., Galanis, D.: Generating natural language descriptions from OWL ontologies: the NaturalOWL system. CoRR **abs/1405.6164** (2014)
3. Third, A., Williams, S., Power, R.: OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies. (2011)
4. Stevens, R., Malone, J., Williams, S., Power, R., Third, A.: Automating generation of textual class definitions from owl to english. J. Biomedical Semantics **2**(S-2) (2011) S5
5. Schiller, M.R.G., Schiller, F., Glimm, B.: Testing the adequacy of automated explanations of EL subsumptions. In: Description Logics. Volume 1879 of CEUR Workshop Proceedings., CEUR-WS.org (2017)
6. Ellampallil Venugopal, V., Kumar, S.P.: A novel approach to generate MCQs from domain ontology: Considering DL semantics and open-world assumption. Web Semantics: Science, Services and Agents on the World Wide Web **34** (2015) 40 – 54
7. Third, A.: "Hidden Semantics": What can we learn from the names in an ontology? In: Proceedings of the Seventh International Natural Language Generation Conference. INLG '12, Stroudsburg, PA, USA, Association for Computational Linguistics (2012) 67–75
8. Grimm, S., Wissmann, J.: Elimination of redundancy in ontologies. In: Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web: Research and Applications - Volume Part I. ESWC'11, Berlin, Heidelberg, Springer-Verlag (2011) 260–274
9. Horrocks, I., Bechhofer, S.: People & Pets ontology. http://owl.man.ac.uk/tutorial/
10. Ellampallil Venugopal, V., Kumar, S.P.: Ontology verbalization using semantic-refinement. CoRR **abs/1610.09964** (2016)