A Human Computation Approach for Ontology Restrictions Verification

Stefani Tsaneva, Marta Sabou

Technical University of Vienna,
Favoritenstrasse 9-11, 1040 Vienna, Austria
stefani.tsaneva@tuwien.ac.at, marta.sabou@tuwien.ac.at

Abstract

Ontologies are semantic resources essential for systems requiring real-world knowledge. As such, their correctness and quality are of high importance, and in some cases can only be achieved with human intervention. In this paper, we propose a Human Computation (HC) solution for the *verification of ontology restrictions* by means of universal and existential quantifiers and report on a controlled experiment to study two core task design aspects: (i) the *formalism to represent ontology axioms* in the HC task and (2) *participant qualification testing*. We find that visual axiom representation and prior knowledge of ontology restriction models lead to best results while prior modeling knowledge reduces the evaluation times. Our findings are of interest to researchers aiming to use HC for knowledge engineering tasks related to ontologies or other conceptual structures (e.g., EER diagrams).

1. Introduction

Ontologies are conceptual models that provide a knowledge representation schema describing concepts of a domain of interest (Kehagias et al. 2008). They are core to many advanced intelligent applications (e.g., chatbots) and an integral component of knowledge graphs. Consequently, ensuring that ontologies are correct and of high quality is crucial, yet it cannot be entirely achieved with automated methods: several quality aspects can only be tested by human involvement (Villalón and Pérez 2016).

One particular verification task that requires a human-in-the-loop is the *evaluation of ontology restrictions* through the usage of universal and existential quantifiers. Several studies (Villalón and Pérez 2016) (Rector et al. 2004) (Warren et al. 2019) have indicated that the use of these quantifiers is not trivial, which leads to ontology defects. For instance, the axiom *"A ProteinLoversPizza is any Pizza that has only Meat toppings"* includes the universal restriction and can be satisfied by the following cases: (a) instances of *ProteinLoversPizza* have one or more *Meat* toppings and no other toppings; (b) instances of *ProteinLoversPizza* have no toppings at all. Often case (b), the trivial satisfaction of the universal restriction, is not intended. This is an issue since modeling the axiom as above results in classifying the *ProteinLoversPizza* as *VegetarianPizza* in a classification system (Rector et al. 2004).

Human Computation & Crowdsourcing (HC&C) is a promising approach to outsource ontology verification tasks to human participants and has already been applied successfully for several such tasks (Sabou et al. 2018). As the verification of ontology restrictions has not yet been addressed with HC, in this paper: (1) we propose an HC-based solution for the verification of ontology quantifiers (Section 2) and (2) we investigate the effects of two essential task design aspects (axiom representation formalism, participant qualification testing) on task performance. Results from a controlled experiments involving junior ontology engineers confirms that visual axiom representation leads to superior performance and that participants with prior modeling knowledge are faster (Sections 3 and 4).

2. Task Design

The task design for verifying ontology restrictions relies on splitting the complex problem of evaluating an ontology into smaller verification tasks focused on a single ontology relation at a time. Prerequisite for the task is the extraction of all existential and universal restrictions from the ontology and their grouping on the same relation forming ontology axioms. Each such axiom represents a small ontology that fully describes a specific relation and can be evaluated independently from the rest of the axioms.

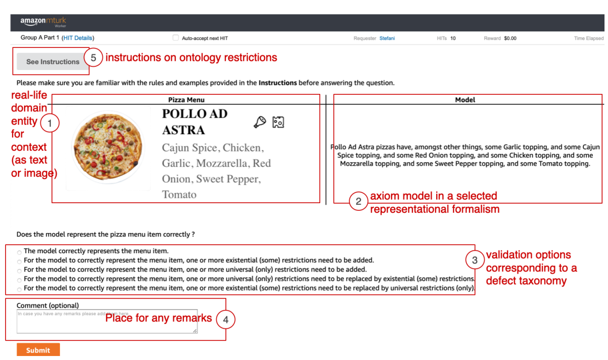


Figure 1: Example of a Human Intelligence Task (HIT) for the verification of ontology restrictions.

Each HC task (Fig. 1) depicts a real-world representation of an entity (1 in Fig. 1) and (2) the corresponding restriction-based ontology axiom (Model). The model can be represented in different formalisms: (a) Rector (Rector et al. 2004) - a textual formalism using *some* and *only* as keywords to represent the restrictions, (b) Warren (Warren et al. 2019) - an alternative paraphrasing using the keywords *at least one* and *no other than* and (c) the graphical representation VOWL[1]. Defect types typical to the usage of the restrictions are identified and organized into a defect taxonomy to be used for guiding workers through the tasks and shifting their focus to possible mistakes. The evaluator's role is to decide whether the axiom model correctly describes the context entity and if not to select a defect type from a number of possible defects (3 in Fig. 1). Workers can also leave free-text comments (4) and inspect relevant instructions (5).

3. Experiment Design

We conducted a controlled experiment, following (Wohlin et al. 2012), to investigate the results which can be achieved with the proposed task design and to test the following hypotheses related to task design:

H1: The formalism in which axioms are represented influences the performance and speed of the contributors.
H2: Prior modeling knowledge has positive influence on the performance and speed of the contributors.

Experimental data was derived from the Pizza Ontology[2], a well known, good quality educational ontology in which we seeded defects manually. In total, our dataset contained 15 incorrect and 15 correct axioms. For the *evaluation population*, we relied on 88 students participating in an introductory lecture on ontology modeling, thus enabling establishing a controlled setting.

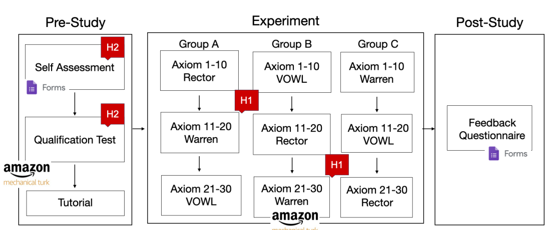


Figure 2: Overview of the experiment workflow.

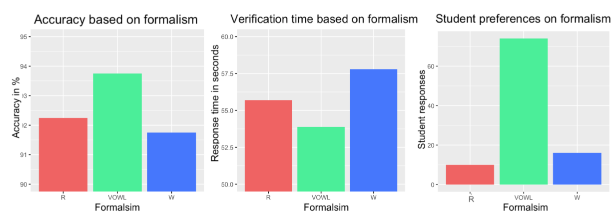
The *experiment setup* (Fig. 2) consisted of three major stages. First, in a *Pre-study* phase student qualification related to ontology modeling topics was assessed both *subjectively* (self-assessment questionnaire) and *objectively* (qualification test). The qualification test consisted in nine increasingly complex modeling tasks and allowed classifying students into three expertise categories (little/some/expert knowledge) based on their performance. This test is one of the contributions of our work, since we are not aware of any previous works that attempted to assess an knowledge engineering expertise level in an objective way. A tutorial on data from a related domain concluded the pre-study phase.

Second, during the *experiment*, students were separated into groups and the data was split into 3 sets. Each group saw the same axiom sets in the same order, however, in a different formalism. For instance, group A started with the first 10 axioms in the Rector formalism, while group B saw those axioms in VOWL. This set-up allowed for a comparison between representations and their influence on the results.

Third, a *feedback questionnaire* concluded the study.

4. Experimental Results

Analysis of the experimental data showed that high accuracy (percentage of correct judgements) of crowdsourced judgments can be achieved using the proposed HC task both for textual and graphical model representations. Related to hypothesis *H1*, Fig. 3 (a) shows for each representational formalism the average percentage of correct responses per HIT and the average verification time per HIT. The verifications performed in the graphical formalism VOWL have slightly higher accuracy (94%) and the average time (54s) needed for evaluating an axiom is lower than in the textual representations. However, the results of a one-way ANOVA test showed no statistical significance of the result differences between the representations. Nevertheless, based on students' feedback, over 70% of the participants preferred working with VOWL (Fig. 3 (b)).



(a) Verification results (b) Contributors preference

Figure 3: Verification results based on the representational formalism of ontology axioms.

Related to hypothesis *H2*, we used Pearson product-moment correlation to investigate the effect of prior modeling knowledge on the verification results. We report that prior knowledge of ontology restriction modeling positively influenced the verification results (correlation coefficient: 0.35; p-value 0.0009353) while prior modeling knowledge reduced the time needed for performing the verification tasks (correlation coefficient: -0.22; p-value: 0.03989).

To conclude, we proposed a HC task design for evaluating ontology restrictions and evaluated it in a controlled experiment showing (1) a high accuracy of results of over 90%; (2) that visual axiom representations are beneficial for performance and preferred for their usability; (3) that it is possible to objectively establish modeling qualification as an indicator of performance. Additionally, our experiment had a positive educational value for the students as a hands-on complement of theoretical lectures. Therefore, we plan a family of experiments associated with our lectures both as an educational tool and a way to investigate additional ontology verification tasks and design alternatives.

[1] http://vowl.visualdataweb.org/v2/
[2] https://protege.stanford.edu/ontologies/pizza/pizza.owl

Acknowledgments

We thank all study participants and FWF for funding this work through the HOnEst project (V 754-N).

References

Kehagias, D. D.; Papadimitriou, I.; Hois, J.; Tzovaras, D.; and Bateman, J. 2008. A methodological approach for ontology evaluation and refinement. In *ASK-IT Final Conference. June.(Cit. on p.)*, 1–13.

Rector, A.; Drummond, N.; Horridge, M.; Rogers, J.; Knublauch, H.; Stevens, R.; Wang, H.; and Wroe, C. 2004. OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In *International Conference on Knowledge Engineering and Knowledge Management*, 63–81. Springer.

Sabou, M.; Aroyo, L.; Bontcheva, K.; Bozzon, A.; and Qarout, R. K. 2018. Semantic Web and Human Computation: The status of an emerging field. *Semantic Web* 9(3): 291–302.

Villalón, M. P.; and Pérez, A. G. 2016. Ontology Evaluation: a pitfall-based approach to ontology diagnosis. *PhD Tesis, Universidad Politecnica de Madrid, Escuela Tecnica Superior de Ingenieros Informaticos* .

Warren, P.; Mulholland, P.; Collins, T.; and Motta, E. 2019. Improving comprehension of knowledge representation languages: A case study with Description Logics. *International Journal of Human-Computer Studies* 122: 145–167.

Wohlin, C.; Runeson, P.; Höst, M.; Ohlsson, M. C.; Regnell, B.; and Wesslén, A. 2012. *Experimentation in software engineering*. Springer Science & Business Media.