netidee

STIPENDIEN

Combining SHACL and Ontologies

Zwischenbericht | Call 18 | Stipendium ID 6794

Inhalt

1. Einleitung

Large amounts of data play more and more an important role in (the background of) modern society. The standardised way to present such data are Knowledge Graphs (KGs), in which some data is only implicitly there: i.e., it can be derived from some basic information and reasoning rules. These reasoning rules are called ontologies. At the same time it is important to be able to check the quality of the data. For this, the SHApes Constraint Language (SHACL) was specifically designed and set as a standard. The goal of this project is to clarify how SHACL can be used together with ontologies, the reasoning rules of KGs, in a meaningful way. Surprisingly, our project is one of the first to tackle this problem, which means that we had to start completely at the beginning and a lot of work remains to be done.

Before the official start of this netidee project, we already developed a theoretical approach for a simplified case of SHACL with a rather inexpressive ontology language. Here the focus was rather technical. In the first months of this project we worked on more practical techniques, optimisations of our main algorithm and expanding the expressivity of the ontology language and the SHACL fragment considered. In other words, we worked both on extending the theory as on the more practical aspects in parallel.

2. Status

2.1. Meilenstein 1 - Accepted paper

We decided to write a paper for the Description Logic Workshop this year in June. We mostly worked on this paper, titled 'Reasoning over the Core Chase: the Case of SHACL Validation over ELHI Knowledge Bases,' in February and March. As it got accepted, this was a useful deviation of the original plan to work on some other projects during that time period.

The problem we are addressing in this paper is the high complexity of the main algorithm we proposed in 'Reconciling SHACL and Ontologies: Semantics and Validation via Rewriting.' Although it is proven that this complexity is needed for the trickiest cases, those cases will most likely not appear in real-live ontologies and SHACL constraints. Thus, we optimised the algorithm, without giving up on generality. Furthermore, we increase the expressivity of the ontology language used, causing some other complexity issues we had to consider and solve.

This paper is a first step towards an implementation of the theory we developed. A goal for some papers in the near feature would be to continue this focus on the practical aspects.

2.2. Meilenstein 2 - Progress article

Next to the paper that focussed on the practical aspects of our project, we also worked on generalising our techniques to a bigger fragment of SHACL and even more expressive ontology languages. In conference papers, there is in general too little space to fully illustrate the technical details of the techniques we are developing, so we are working on a journal article. We made a lot of progress on this article: almost all technical details are worked out and we can slowly start focussing on finishing up.

2.3. Meilenstein 3 - International collaboration

In the first version of the project schedule, just a one week research visit to the TU Dresden was planned. Thus, I went there in February, gave a seminar and had many interesting research meetings with some researchers there. As this visit turned out to be so relevant, we planned another, extra and longer visit for this year: for May until mid July.

In Dresden, there is currently a research project going on in which an ontology and a set of SHACL constraints to perform risk analysis was developed. This project also contains an implementation by first performing the ontology reasoning (which is possible because of the type of ontology language, but not possible in general) followed by a separate application of an already existing SHACL validator. One of the projects I will be working on here is using the techniques we developed before to encode the ontology reasoning in the SHACL validator, causing the ontology reasoner to become superfluous.

3. Zusammenfassung Planaktualisierung

The following points changed in the schedule:

- While working on one of the papers, our attention shifted to some other projects that are more pressing now. The results already obtained can be updated to a paper in the future: possibly next autumn or winter.
- We wrote an extra paper not specified in the previous schedule for the Description Logic Workshop, which got accepted, see also milestone 1. This caused some delay in finishing the article we were writing: see milestone 2 for more on this article.
- After my one week research visit to TU Dresden in February, we decided that a longer visit there would be fruitful: this longer visit is currently happening, see also milestone 3, and had some influence on the schedule.
- We decided on which conferences and summer schools I am going to attend this summer and what our next focus, research-wise, should be.
- Some extra given seminar talks and an extra workshop visit were added.