netidee STIPENDIEN

Combining SHACL and Ontologies

Final Report | Call 18 | Stipend ID 6794

Inhalt

1. Introduction

As our society relies more and more on collecting and analysing huge amounts of data in many different formats, the need for efficient and flexible tools to handle this data heavily increases. The *Resource Description Framework* (RDF) is a graph-based model for managing data on the Web, introduced by the World Wide Web Consortium (W3C). Currently, it is the standard way how web data is published and shared. It is for instance used in sharing health care information: Fast Healthcare Interoperability Resources (FHIR), helps in deciding the best possible treatment for patients. Another use case is assistance in failure analysis for semi-conductors.

In parallel with RDF, the W3C introduced the *Ontology Web Language* (OWL) to reason about implicit knowledge, like 'each student has a student number'. This information is captured in so-called ontological axioms, which are used for reasoning over and completion of the schematised knowledge. The profiles of OWL are based on *Description Logics* (DLs).

A demand surfaced for a language to describe and validate the correctness of RDF graphs. OWL is not designed for this, as the axioms only state that we know the information exists, while you might want to check availability of certain data. I.e., you might want to check that each student has indeed a student number assigned instead of reasoning over the fact that we know that each student must have one. This need to test the quality of huge amounts of data led to the introduction of SHACL.

Although lots of issues concerning SHACL remain open and aspects are lacking, the basics of SHACL are already widely adopted to detect irregularities in data as it is so important in practice. However, to have reliable SHACL technologies, it is much needed to understand SHACL better and provide solid foundations. To summarise, plain SHACL as proposed in the SHACL standard is already in need of technical guidance.

On top of that, the SHACL standard mentions SHACL validation combined with ontological knowledge as an important feature. However, without any guidance how to do so. Our project focuses on providing solid foundations for this combination. This is a challenging problem, as it involves designing an intuitive meaning for combining inferred knowledge (OWL) with constraint checking (SHACL). Although there is great interest in the combination, prior to this project no satisfying proposals have been put forth.

2. Research questions

The main topic of this research project can be summarised in one research question.

- **What does it mean to combine SHACL and Ontologies?**

More concretely, we break this down into four concrete subquestions.

1. What are adequate semantics for validating SHACL in presence of ontologies?

This means finding a meaningful combination of SHACL and different description logics, and aligning the technical details of both. For the smaller fragments, the main difficulty lies in finding an intuitive semantics. For more expressive fragments of SHACL and description logics the main challenge will be a computational one: whether it is possible and how to include the whole set of features of SHACL in a meaningful way, while preserving decidability.

2. What is the complexity of these semantics?

The next step is to develop optimal algorithms for determining validation of SHACL combined with ontologies, to determine upper bounds of the complexity for the different fragments we are considering. For determining the lower bounds we can rely on straightforward methods: finding the right problem in a certain complexity class and reformulating this problem in our setting.

3. What is the implementability of these semantics?

Can we get to practical techniques? It seems that the most promising path is reducing validation of SHACL with ontologies to plain validation of SHACL, by rewriting the constraints based on ontological reasoning. However, it seems that these algorithms will need quite some computational complexity power to take care of all the technical corner cases. Thus, we most likely will have to put some restrictions on the SHACL constraints, and possibly also on the ontology language, to find algorithms for answering whether a certain given shapes graph is validated by a given ontology on some RDF graph, and such that these algorithms are at most polynomial in size of the combined complexity of the rules in the ontology and the RDF graph. Deciding which restrictions are the most meaningful, i.e., are mostly excluding some theoretical, complicated, constructions and not the main features used, is also part of solving this subquestion.

4. Can we find decidable instances of SHACL satisfiability and containment in presence of ontologies?

SHACL satisfiability asks the question whether for a given set of constraints and targets there exists a data graph that validates the targets. The

containment problem requires two sets of constraints and targets and asks whether all graphs that validate the targets of the first set, also validate the targets of the second set. This is a more static analysis of the combination. In practice, answering this question helps to find more efficient SHACL constraints and helps us understand what we are asking exactly with our constraints.

3. Results

During the term of this stipend, we have written and published the following paper:

> Anouk M. Oudshoorn, Magdalena Ortiz, Mantas Šimkus, *Reasoning with the Core Chase: the Case of SHACL Validation over ELHI Knowledge Bases.* **Description Logics 2024,** Bergen, Norway.

This paper addresses how to generalise the results in [2] to a more complex ontology language, the description logic ELHI, and at the same time makes a first step towards implementability. That is, instead of considering all possible types, we developed a smart system that given some databases or set of starting types, determines which subset of types suffice to get a sound and complete rewriting algorithm, that reduces SHACL and OWL validation to plain SHACL validation.

Next to this paper, we have written a journal article that is ready for submission, but not yet published. In this article, we extend our results even further, to a normalised version of the again more expressive ontology language Horn-ALCHIQ. That is, we provide a semantics and rewriting to make this semantics more feasible. Moreover, among other results, we establish a normal form for recursive SHACL, and determine the complexity of deciding validation of SHACL in presence of the considered ontology language. Moreover, we have written another paper that is currently under double-blind submission for presentation at an international conference.

4. Planned follow-up activities

There are multiple open projects continuing the results achieved. One of them consists of working on a concrete implementation for yet another fragment, whereas another is lifting our developed techniques to the more general case of existential rules (instead of OWL), over which we consider a certain query language (instead of SHACL). These projects are the result of my previous research stay at the Technical University of Dresden, which took place during the term of this stipend, and have seen great progress.

Recently, we also started looking into the last research question posed: can we decide satisfiability for a certain fragment of SHACL, possibly in combination with OWL.

To finish my dissertation, there are at least another 18 months left to complete any leftover projects and answer the proposed questions to a larger extend.

5. Suggestions for continuations by third parties

We strongly believe in the usefulness of combining SHACL and OWL and performed preliminary work in this regard. However many fragments remain untouched from only the theoretical perspective already, resulting in a whole bunch of questions that are left to be answered. Furthermore, when moving towards implementations or practical tools, there is even more work left to do.

Moreover, we developed some nice techniques that need not be limited to this project. In fact, we have already encountered some examples, like the generalisation to existential rules, in which our techniques appear to be useful outside of the initial setting we developed them for, and there must be many more such cases we are not (yet) aware of.

6. Bibliography

[1] S. Abiteboul, R. Hull, and V. Vianu, *Foundations of Databases,* Addison-Wesley, 1995.

[2] S. Ahmetaj, M. Ortiz, A.M. Oudshoorn, and M. Šimkus, *Reconciling SHACL and Ontologies: Semantics and Validation via Rewriting,* in Proc. of the 26th ECAI, p.27-35. 2023.

[3] M. Andresel, J. Corman, M. Ortiz, J.L. Reutter, O. Savkovic, and M. Šimkus, *Stable Model Semantics for Recursive SHACL*, in Proc. of The Web Conference 2020, p.1570-1580. ACM, 2020.

[4] K.R. Apt, H.A. Blair, and A. Walker, *Towards a Theory of Declarative Knowledge,* p.89-148, Morgan Kaufmann Publishers, 1988.

[5] M. Arenas, G. Gottlob, and A. Pieris, *Expressive Languages for Querying the Semantic Web,* ACM Trans. Database Syst. 43, p.13:1-13:45. 2018.

[6] F. Baader, S. Brandt, and C. Lutz, *Pushing the EL Envelope,* in Proc. of the 19th IJCAI, p.364-369. Professional Book Center, 2005.

[7] F. Baader, I. Horrocks, C. Lutz, and U. Sattler, *An Introduction to Description Logics,* Cambridge University Press, 2017.

[8] F. Baader, C. Lutz, and S. Brandt, *Pushing the EL Envelope Further,* in Proc. of the 4th OWLED. CEUR-WS.org, 2008.

[9] M. Bienvenu, M. Ortiz, *Ontology-mediated Query Answering with Data-tractable Description Logics,* in Reasoning Web, volume 9203 of LNCS, p.218-307. Springer, 2015.

[10] B. Bogaerts, M. Jakubowski, and J. Van den Bussche, *SHACL: A Description Logic in Disquise,* in Proc. of LPNMR, LNCS 13416, p.75-88. Springer, 2022.

[11] S. Borgwardt and W. Forkel, *Closed-world semantics for conjunctive queries with negation over ELH-bottom ontologies*, in Proc. of the 28th IJCAI, p.6131-6135. ijcai.org, 2019.

[12] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriquez-Muro, and G. Xiao, *Ontop: Answering SPARQL Queries over Relational Databases,* Semantic Web 8, p.471-487. 2017.

[13] D. Calvanese, G.D. Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriquez-Muro, R. Rosati, and D.F. Savo, *The MASTRO System for Ontology-based Data Access,* Semantic Web 2, p.43-53. 2011.

[14] D. Carral, M. Krötzsch, M. Marx, A. Ozaki, and S. Rudolph, *Preserving Constraints with the Stable Chase,* in Proc. of 21st. ICDT, p.12:1-12:19. Schloss Dagstuhl, 2018.

[15] A. Chmurovic and M. Simkus, *Well-founded Semantics for Recursive SHACL,* in Proc. of the 4th Datalog-2.0, p.2-13. CEUR-WS.org, 2022.

[16] J. Corman, J.L. Reutter, and O. Savkovic, *Semantics and Validation of Recursive SHACL,* in Proc. of the 17th ISWC, p.318-336. Springer, 2018.

[17] A. Deutsch, A. Nash, and J.B. Remmel, *The Chase Revisited,* in Proc. of the 27th ACM-PODS, p149-158. ACM, 2008.

[18] T. Eiter, M. Ortiz, M. Šimkus, T.-K. Tran, and G. Xiao, *Query Rewriting for Horn-SHIQ plus rules,* in Proc. of the 26th AAAI, p.726-733. AAAI Press, 2012.

[19] S. Ellmauthaler, M. Krötzsch, S. Mennicke, *Anwering Queries with Negation over Existential Rules,* in Proc. of 26th AAAI, p.5626-2633. AAAI Press, 2022.

[20] R. Fagin, P.G. Kolaitis, R.J. Miller, and L. Popa, *Data Exchange: Semantics and Query Answering,* Theor. Comput. Sci. 336, p.89-124. 2005.

[21] W. Fan and F. Geerts, *Relative Information Completeness,* ACM Trans. Database Syst. 2010.

[22] P. Hell and J. Nesetril, *The Core of a Graph,* Discret. Math. 109, p.117-126. 1992.

[23] J. Hromkovic, S. Seibert, and T. Wilke, *Translating Regular Expressions into small e-free Nondeterministic Finite Automata,* J. Comput. Syst. Sci. 62, p.565-588. 2001.

[24] M. Knorr, J.J. Alferes, and P. Hitzler, *Local Closed World Reasoning with Description Logics under the Well-founded Semantics,* Artif. Intell. 2011.

[25] H. Knublauch, D. Allemang, and S. Steyskal, *SHACL Advanced Features,* W3C Working group note. W3C, 2017.

[26] H. Knublauch and D. Kontokostas, *Shape Constraint Language (SHACL),* W3C Recommendation. W3C, 2017.

[27] M. Krötzsch, *OWL 2 Profiles: An Introduction to Lightweight Ontology Languages,* in Reasoning Web - 8th International Summer School, p. 112-183. Springer, 2012.

[28] M. Krötzsch, *Computing Cores for Existential Rules with the Standard Chase and ASP,* in Proc. of the 17th KR, p.603-613. 2020.

[29] D.S. Johnson, and A.C. Klug, *Testing Containment of Conjunctive Queries under Functional and Inclusion Dependencies,* Journal of Comput. Syst. Sci. 28, p.1039-1044. 1984.

[30] B. Marnette, *Generalized Schema-mappings: from Termination to Tractability,* in Proc. of the 28th PODS, p.13-22. ACM, 2009.

[31] B. Motik, I. Horrocks, and U. Sattler, *Adding Integrity Constraints to OWL,* in Proc. of OWLED. CEUR-WS.org, 2007.

[32] B. Motik, I. Horrocks, and U. Sattler, *Bridging the Gap between OWL and Relational Databases,* in Proc. of WWW. ACM, 2007.

[33] M. Ortiz, S. Rudolph, and M. Šimkus, *Query Answering in the Horn Fragments of the Description Logics SHOIQ and SROIQ,* in Proc. of the 22th IJCAI, p.1039-1044. IJCAI/AAAI, 2011.

[34] M. Ortiz and M. Šimkus, *Reasoning and Query Answering in Description Logics,* in Reasoning Web, volume 7487 of LNCS, p. 1-53. Springer, 2012.

[35] A.M. Oudshoorn, M. Ortiz, and M. Šimkus, *Reasoning with the Core Chase: the Case of SHACL Validation over ELHI Knowledge Bases,* in Proc. of the 37th. DL. CEUR-WS.org, 2024.

[36] P. Pareti, G. Konstantinidis, and F. Mogavero, *Satisfiability and Containment of Recursive SHACL,* Journal of Web Semantics, 2022.

[37] H. Pérez-Urbina, B. Motik, and I. Horrocks, *Tractable Query Answering and Rewriting under Description Logic Constraints,* J. Appl. Log. 8, p.186-209. 2010.

[38] O. Savkovic, E. Kharlamov, and S. Lamparter, *Validation of SHACL Constraints over KGs with OWL 2 QL Ontologies via Rewriting,* in Proc. of the 16th ESWC, p.214-329. Springer, 2019.