Validating RDF data using Shapes

Jose Emilio Labra Gayo[a]

[a] *University of Oviedo, Spain*

Abstract

RDF forms the keystone of the Semantic Web as it enables a simple and powerful knowledge representation graph based data model that can also facilitate integration between heterogeneous sources of information. RDF based applications are usually accompanied with SPARQL stores which enable to efficiently manage and query RDF data. In spite of its well known benefits, the adoption of RDF in practice by web programmers is still lacking and SPARQL stores are usually deployed without proper documentation and quality assurance. However, the producers of RDF data usually know the implicit schema of the data they are generating, but they don't do it traditionally.

In the last years, two technologies have been developed to describe and validate RDF content using the term shape: Shape Expressions (ShEx) and Shapes Constraint Language (SHACL). We will present a motivation for their appearance and compare them, as well as some applications and tools that have been developed.

Keywords

RDF, ShEx, SHACL, Validating, Data quality, Semantic web

1. Introduction

RDF is a flexible knowledge representation language based of graphs which has been successfully adopted in semantic web applications. In this tutorial we will describe two languages that have recently been proposed for RDF validation: Shape Expressions (ShEx) and Shapes Constraint Language (SHACL).ShEx was proposed as a concise and intuitive language for describing RDF data in 2014 [1]. The syntax of ShEx is inspired by SPARQL. ShEx has been recently adopted in several projects like Wikidata [2]. SHACL was accepted as a W3C recommendation in 2017[1] and has also been adopted by a large number of companies.

Although both ShEx and SHACL have similar goals, the underlying philosophy is different: while ShEx schemas provide descriptions about the expected RDF data, SHACL shapes graphs provide constraints and things that are not allowed [3].

In the tutorial we will present an overview of both and describe some challenges and future work [4]

2. Acknowledgements

This work has been partially funded by the Spanish Ministry of Economy, Industry and Competitiveness, project: TIN2017-88877-R

3. References

[1] Prud'hommeaux, E. , Labra Gayo, J. E., Solbrig, H.: Shape expressions: an RDF validation and transformation language. SEMANTICS 2014: 32-40

[2] Thornton, K., Solbrig, H., Stupp, G. S., Labra Gayo, J.E., Mietchen D., Prud'hommeaux, E., Waagmeester, A.: Using Shape Expressions (ShEx) to Share RDF Data Models and to Guide Curation with Rigorous Validation. ESWC 2019: 606-620

[3] José Emilio Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, Dimitris Kontokostas: Validating RDF Data., Morgan & Claypool Publishers, 2017

ISIC'21:International Semantic Intelligence Conference, February 25–27, 2021, New Delhi, India

✉ : labra@uniovi.es (Labra Gayo)

[1] https://www.w3.org/TR/shacl/

[4] Labra Gayo, J.E., García-González H., Fernández-Alvarez D., Prud'hommeaux E. (2019) Challenges in RDF Validation. Studies in Computational Intelligence, vol 815. Springer, Cham.