

Incremental Validation of ShEx for RDF

Keywords: RDF, Shape Expression Schemas, Algorithms

Supervisors: Iovka Boneva (@univ-lille.fr)

Location: Links project-team of Inria and CRISTAL, Haute Borne

Context:

RDF is the data format recommended by the Semantic Web and the Linked Data initiatives for publishing data by institutions and organisations. The result is a constantly growing collection of interconnected datasets. Some are freely available and used in e.g. biology, medical and pharmaceutical research, cultural heritage databases, data journalism. Other are proprietary, as for instance news archives.

The Shape Expressions Schemas (ShEx) language allows to assess the quality of RDF data with respect to the requirements of a particular application. This is done by defining constraints and the data is called valid when it satisfies the constraints. None of the existing ShEx software allows to take into account the dynamic aspect of RDF datasets: when the data evolves, all the constraints need to be checked from scratch.

Tasks:

The project focuses on the incremental validation of ShEx, including these tasks:

- **Identify how changes in the data influence its validity.** When facts are added or removed, the dataset might become invalid. The objective is to classify the different ways in which updates of the data can break its validity.
- **Devise an efficient incremental validation algorithm.** An incremental algorithm stores intermediate results of computation. When the algorithm is called for checking the validity of the updated dataset, it can use the stored results and save computation time in the cost of using additional space. An efficient algorithm should allow to find a good trade-off between additional space consumption versus time gain.
- **Implement and test the algorithm.** The efficiency of the incremental algorithm is to be tested on big datasets.

Bibliography :

- **Shape Expressions Language 2.0.** <http://shex.io/shex-semantics/>. E. Prud'hommeaux, I. Boneva, J. Labra Gayo, G. Kellogg.
- **Semantics and Validation of Shapes Schemas for RDF.** In ISWC2017 - 16th International semantic web conference. I. Boneva, J. Labra Gayo and E. Prud'hommeaux.
- **RDF :** <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html>
- **Linked Data.** <http://linkeddata.org/>. Contains references to public linked datasets.