

Tips to improve Semestral work

Miroslav Blaško

November 21, 2019

1 Related to report

Consider to describe transformation using data-flow diagram in cases where it is too complex to describe by text.

2 Related to data

Note that not all tips apply to every situation and its feasibility should be evaluated w.r.t. you project.

1. **provide data provenance** – In general we are trying to integrate datasets and thus datasets should be not transformed too much in CP1. Moreover, from output RDF it should be possible to track the origin of data (e.g. we want to know that this RDF triple came from a concrete CSV and its concrete column. One of the ways to do it is following: we will use for each processed CSV separate RDF namespace and for each RDF property we provide metadata from which *column* name it originates.

2. **obey RDF/RDFS/OWL specification**

- a) **URI should represent real things** – Triples should be possible to read as sentences. Bad example would be:

```
:row123 a :BuildingId .  
:row123 :capacity 12 .
```

Building has capacity, not BuildingId has capacity. Moreover, *row123 has capacity 12* does not make sense as sentence as well.

- b) **use datatypes correctly** – Bad example:

```
□:x :hasYear "2019.0"^^xsd:double .
```

- c) **use RDF/RDFS/OWL terms correctly** – i.e. differentiate *owl:Classes* vs. *owl:NamedIndividual*, *owl:ObjectProperty* vs. *owl:DataProperty*, specify domain and range of properties.

3. **use appropriate file extensions** – SPARQL queries (*.sparql, *.rq), RDF/RDFS/OWL files in RDF/XML format (e.g. *.rdf), in Turtle format use *.ttl.
4. **use specific IRI instead of example.org** – e.g. `http://onto.fel.cvut.cz/ontologies/os`