

In this thesis we study different techniques for expressing and verifying integrity constraints over data stored in ontologies built using Semantic Web technologies.

This work is motivated by data validation requirements in the @Web platform, a semantic web application that allows domain experts to extract data in scientific documents, and researchers to explore and query those data via a graphical user interface. The extracted data is stored in a publicly accessible RDF graph following a predefined OWL ontology, and shared with the research community.

Given the error-prone nature of the data extraction process, a set of integrity constraints has been identified that all extracted data must fulfill. It is desired to validate these constraints automatically and report any validation errors to the domain expert during the data extraction process.

To this end, we first survey the current W3C recommendations for querying, describing and constraining the contents of RDF graphs and the available tools implementing these recommendations. We decide to focus our analysis on SPARQL, Shape Expressions and SHACL. We then implement a set of test constraints using each of the available tools and compare them according to expressiveness, verbosity, readability, running times, etc. Finally, we identify the tool that best suits our concrete needs and proceed to implement the constraint validation features in @Web.

Our analysis shows that plain SPARQL queries yield the best running times amongst the technologies considered. We also observe that certain constraints expressed in both Shape Expressions and SHACL require nesting SPARQL queries that are comparable in length to stand-alone SPARQL queries implementing those same constraints, thus defeating the purpose of an alternate constraint language. To conclude, we propose modifications to both the Shape Expressions and SHACL languages that would allow all our use cases.

**Keywords:** Semantic Web, RDF, OWL, SPARQL, Shape Expressions, SHACL, Constraints, @Web.