Knowledge Lab

Birkbeck UNIVERSITY OF LONDON

Research Student
Riccardo Frosini

Supervisors
Alexandra Poulovassilis
Andrea Calì
Peter T. Wood

Flexible Query Processing of SPARQL Queries

Research Aims

Our research aim is to extend **SPARQL 1.1** with two operators; **APPROX** and **RELAX**, previously introduced in the context of regular path queries. Using these operators we are able to support flexible querying over the property path queries of SPARQL 1.1. We call this new language **SPARQL^{AR}**. Flexible querying techniques have the potential to enhance users' access to complex, heterogeneous datasets, by allowing the retrieval of non-exact answers to queries that are related in some way to the exact answers. Users are able to query RDF data without fully knowing the structure of a dataset. APPROX and RELAX encapsulate different aspects of query flexibility namely, finding different answers and finding more answers, respectively.

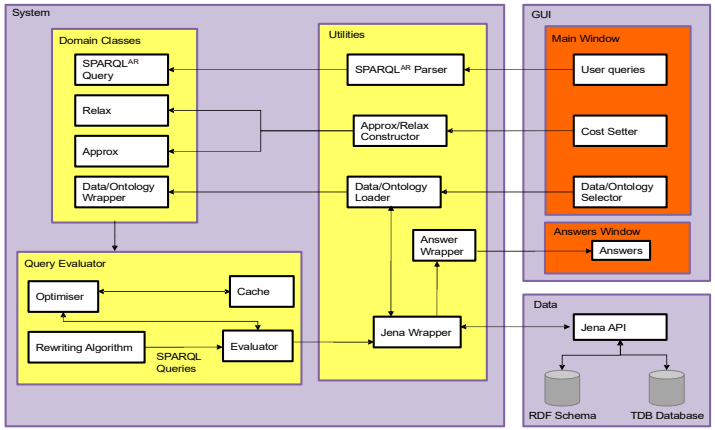

Figure 1. Software Architecture

Research Methodology

The focus of this study is to formalise the semantics of SPARQL^{AR}, investigate its complexity and devise an evaluation algorithm based on query rewriting (which generates multiple SPARQL 1.1 queries by rewriting a SPARQL^{AR}). We also propose three optimisation techniques for increasing the performance of the evaluation. The first optimisation is a pre-computation technique that caches the answers of parts of the queries generated by the rewriting algorithm. These answers will then be reused to avoid the re-execution of those sub-queries. The second optimisation utilises a summary of the dataset to discard queries that it is known will not return any answer. The third optimisation technique uses the query containment concept to discard queries whose answers would be returned by another query. We have implemented SPARQL^{AR} query evaluation and all three optimisation techniques (Figure 1).

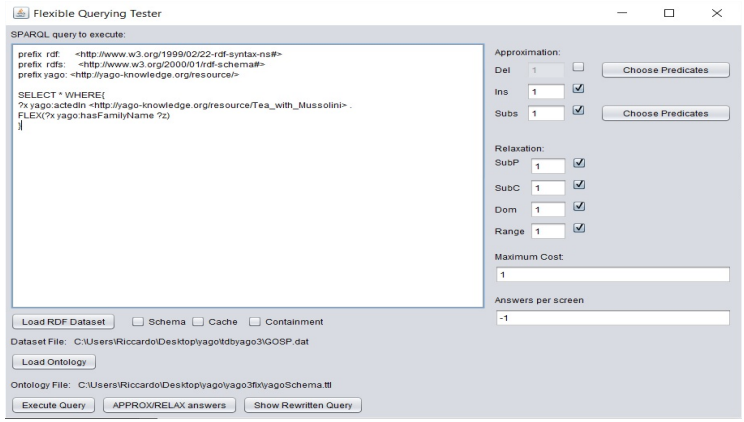

Figure 2. User Interface

Performance Study

By the means of the prototype and a UI tool (see Figure 2) for evaluating SPARQL^{AR} queries, we have undertaken performance evaluation by running multiple SPARQL^{AR} queries over three datasets, namely: **DBPedia**, **LUBM** and **YAGO**. We have shown the benefit of using the optimisation techniques proposed in terms of query execution time. Currently, we are undertaking a more extensive evaluation study by evaluating SPARQL^{AR} queries using all three optimisation techniques combined.

Publications

Riccardo Frosini, Andrea Calì, Alexandra Poulovassilis, Peter T. Wood: **Flexible Query Processing for SPARQL.** Semantic Web Journal, 8(4), pp 533-563 (2017)

Riccardo Frosini, Andrea Calì, Alexandra Poulovassilis, Peter T. Wood: **Flexible Querying for SPARQL.** OTM Conference, pp 473-490 (2014)

Birkbeck UNIVERSITY OF LONDON

Department of Computer Science and Information Systems