Worst-case Optimal Join Algorithms for SPARQL Query Evaluation

Master Thesis Proposal

The Resource Description Framework (RDF) [1] has become the de facto data model for representing information in the Web. Data in RDF are modeled as triples of the form <subject predicate object> where subjects and objects stand for resources of any kind whereas predicates denote relationships between resources. Given that the same resource can be subject and object in different triples, an RDF dataset forms a directed labeled graph like the one depicted in Fig 1a. The default language for querying RDF graphs is SPARQL [2] with a declarative syntax similar to that of SQL. Queries in SPARQL define graph patterns that have to be evaluated on the RDF graph; for instance, the SPARQL query in Fig. 1b asks for a pair of married persons who were born and died in the same city. Current state-of-the-art RDF stores evaluate SPARQL queries either with pair-wise join algorithms from databases or with graph exploration techniques. However, none of the existing SPARQL engines has been compared with the recently introduced worst-case optimal join algorithms [3].

![Fig. 1: An Example RDF Graph with a SPARQL Query](image)

The proposed master thesis will focus on the application of worst-case optimal join algorithms for in-memory SPARQL query evaluation. The analysis will be based on the Rust implementation of Differential Dataflow [4], a novel data-parallel programming framework that offers an efficient version of the NPRR algorithm from [3]. The goal of the thesis is twofold. First, we will investigate the performance of worst-case optimal join algorithms for various SPARQL benchmarks [5]. Second, we will compare our results with the state-of-the-art SPARQL query engines that use pair-wise join algorithms and graph exploration techniques.

If you are interested in this project and want to discuss further, please contact John Liagouris (liagos@inf.ethz.ch) and Desislava Dimitrova (desislava.dimitrova@inf.ethz.ch). The proposed Master Thesis will be supervised by Prof. Timothy Roscoe.

[5] https://www.w3.org/wiki/RdfStoreBenchmarking