

## Graduiertenkolloquium Angewandte Informatik

# Decentralized Query Processing over Heterogeneous Sources of Knowledge Graphs

**M.Sc. Lars Heling**

Knowledge Graphs have recently gained attention as a knowledge representation technique for scientific and industrial applications. Knowledge Graphs describe a set of typed entities with their attributes and the relationships between those entities. Linked Data is a popular framework for representing and publishing Knowledge Graphs on the Web. The foundations of Linked Data are the Resource Description Framework (RDF) as a graph-based data model and SPARQL as the query language for RDF. The increasing number and size of Knowledge Graphs published as Linked Data led to the development of various interfaces to support querying Knowledge Graphs on the Web. These interfaces, ranging from Triple Pattern Fragment servers to SPARQL endpoints, are mainly characterized by their emphasis on querying availability and expressivity. The Linked Data Fragment Framework provides a uniform way to describe these interfaces regarding those characteristics.

Decentralized management and heterogeneous interfaces used to publish Knowledge Graphs leads to new challenges for client-side SPARQL query processing. First, many traditional query planning approaches rely on fine-grained statistics on both the interface's querying performance and the data distributions of the Knowledge Graphs. Therefore, performance and data distribution profiling approaches need to be adapted to the capabilities and limitations of different Linked Data Fragment interfaces. Second, in case such statistics are not available, query planning approaches should still be able to obtain efficient query plans that are robust with respect to potential errors during query planning. Finally, when querying heterogeneous federations of Knowledge Graphs, query planning approaches need to be aware of and leverage the querying capabilities of the interfaces present in the federation to reduce both the load on the servers as well as the query execution times. In this talk, we present two approaches for efficient query processing over heterogeneous sources of Knowledge Graphs that address the final two challenges. The first approach focuses on cost- and robustness-based query planning to devise efficient query plans over Triple Pattern Fragment servers. To this end, we propose a cost-model for query plans and introduce a concept of robustness for SPARQL query plans, reflecting the impact of cardinality estimation errors on the cost. In our evaluation, the proposed approach outperforms existing state-of-the-art heuristic-based optimizers validating the effectiveness of combining both cost and robustness. In the second approach, we focus on heterogeneous federations of Linked Data Fragment interfaces. We propose a framework for SPARQL query processing over such heterogeneous federations that is comprised of interface-aware query decomposition, query planning, and polymorphic join operators. The results of our evaluation show that leveraging the capabilities of different Linked Data Fragment interfaces in a federation allows for reducing both the execution time and the number of requests for evaluating query plans.

**Termin: Freitag, 22. Januar 2021, 14:00Uhr**

Ort: Onlineveranstaltung

Zoom-Link: <https://kit-lecture.zoom.us/j/4136684547?pwd=dTNrZmpmRWxoY2NNWxcxV2RDT3Vqdz09>

Meeting ID: 413 668 4547

Passcode: #AaE67JJji

Veranstalter: Institut AIFB, Forschungsgruppe Web Science

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

A. Oberweis, H. Sack, A. Sunyaev, Y. Sure-Vetter (Org.), M. Volkamer, J. M. Zöllner