

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Marzia Cutajar
Název práce Querying RDF data in Multi-Model NoSQL databases
Rok odevzdání 2020
Studijní program Informatika **Studijní obor** Softwarové a datové inženýrství

Autor posudku Martin Svoboda **Role** Vedoucí
Pracoviště KSI

Text posudku:

Based on the analysis of the existing approaches, the aim of this thesis was to propose a new approach that would be dealing with storing and querying of RDF data in the context of multi-model or core NoSQL database systems. These systems more or less successfully tackle not just the variety characteristic of the contemporary data, yet do not directly and natively support RDF semantic data model at the logical level.

In particular, the author chose to work with ArangoDB, one of the existing multi-model representatives, supporting namely the JSON document model as the native one. First, two new and improved ways of transforming RDF data into a form of ArangoDB documents, edge documents, and collections were proposed. The core part of the work then consists of an approach of transforming SPARQL queries into the corresponding ArangoDB queries. It means that an input SPARQL query expression is first parsed into a SPARQL algebra tree (using a third-party library), this tree then transformed into an internal AQL query tree representation, so that this tree is serialized into a target AQL query expression in the end.

The thesis itself is written in high-quality English without grammar mistakes or other shortcomings. It has the expected extent as for the number of pages, as well as it contains all the usually involved components. The proposed transformation framework is introduced in a formal way and also illustrated on particular examples, yet certain parts of the thesis could have been enriched with explanatory paragraph texts and more elaborately prepared examples, too. Using the provided prototype implementation, sample evaluation experiments on real-world data were conducted. Although their results may not be convincing enough at first, it is essential to realize that NoSQL systems, in general, are based on the horizontal scaling approach, and so their true benefit only appears when working with Big Data.

To conclude, Marzia Cutajar worked on the thesis intensively, attended regular meetings as well as proved the required abilities and knowledge of the area. The thesis assignment itself was fulfilled in all aspects. Therefore, I recommend this thesis to be defended successfully.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Datum 26. ledna 2020

Podpis