1. Motivation

The maintenance and use of (statement-level) metadata, such as provenance or time-related information (when a subject created or retrieved), is of increasing importance in the Semantic Web. Especially for Big Data applications, that work on heterogeneous data fused from multiple sources and which require high data quality, data processing metadata is essential. The storage of such metadata alongside the data in the same RDF store allows to record fine-grained traceability and provenance information, license and access rights, data trustworthiness and confidence scores for every single fact in a knowledge graph. However, studies investigating the performance of Metadata Representation Models (MRMs) show that choosing the appropriate metadata representation depends on the used data and metadata, queries and RDF store. Thus it is challenging to determine the best MRM for a scenario beforehand. To enable the development of an RDF application extending metadata use, but without restricting it to a priori one concrete implementation of an MRM, we propose MaSQue (Metadata Storage and Querying).

4.1 Meta-RDF: JSON-based metadata representation format

MaSQue is a Java-based framework and command line utility. Its paradigm is to hide the complexity and individual characteristics of various Metadata Representation Models (MRMs) behind a uniform "mask". It allows to switch between different MRMs without rewriting the application logic. It consists of 2 major components meta-RDF and meta-SPARQL, which establish an abstraction layer for RDF data and its metadata for storage & serialization and querying respectively.

2. Metadata Representation Models (MRMs)

Six different ways of describing (or reifying) an RDF triple $s$, $p$, $o$ with a metadata key and value pair are supported by MaSQue: Companion property (cpprop), nary relation (nary), named graphs (ngraph), singleton properties (sprop), standard reification (strefl), and the Blazegraph-specific Reification Done Right (rdr). Besides rdr, which is based on the vendor-independent RDF* and SPARQL*, all approaches use an explicit statement identifier (id), which is used to attach metadata (green) to the data (grey).

3. MaSQue Approach

MaSQue is a Java-based framework and command line utility. Its paradigm is to hide the complexity and individual characteristics of various Metadata Representation Models (MRMs) behind a uniform "mask". It allows to switch between different MRMs without rewriting the application logic. It consists of 2 major components meta-RDF and meta-SPARQL, which establish an abstraction layer for RDF data and its metadata for storage & serialization and querying respectively.

MaSQue Dataflow

4.2 Meta-RDF Data Model (DAO)

While the JSON format is intended for a batch conversion of a complete dataset, applications can also use the integrated Java data model abstraction (DAO) to convert RDF metadata on-the-fly.

5. Meta-SPARQL: Annotated SPARQL Query Translation

In order to enable MRM-independent SPARQL queries, the generic and extensible tool meta-SPARQL has been developed. It allows automatic rewriting of SPARQL queries for different MRMs. The idea is to, replace every triple pattern within a SPARQL query by a set of special annotations, which will be translated by meta-SPARQL into the appropriate format. Every query needs to be written as a template in an intermediate SPARQL dialect based on these annotations. It consists of 4 annotations explained in the Table on the left. The template can be converted into query instances of the various MRMs. Therefore query templates can be written independent of granularity support and other MRM-specific characteristics.

License:

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)}

ACKNOWLEDGMENTS

This work was supported by grants from the Federal Ministry for Economic Affairs and Energy of Germany (BMWi) for the Smart Data Web project (SA-0144010G18), as well as from the European Union for the Horizon 2020 project ALIGNED (GA 644459). Special thanks go to Markus Helfrich for his inspiring support. Printed at Universitätsdruckerei Leipzig.

Links:

https://papersofweb.com/2019/06/01
https://web.archive.org/web/201908/01

CONTACT

Johannes Frey
jfrey@informatik.uni-leipzig.de
Sebastian Hellmann
hellmann@informatik.uni-leipzig.de
Knowledge Integration and Linked Data Technologies (AKSW/KILT), research group
Leipzig University & Institute for Applied Informatics (IVIA), Leipzig, Germany

HOSTED BY

Smart Data Web Laboratory

SUPPORTED BY

InD4, Leipzig, Germany

SA 4.0)