Overview of Graph Database Solutions

Ernesto Jiménez-Ruiz
Lecturer in Artificial Intelligence

Agenda for today

- Reminder: module evaluation (30% so far)
- MSc project at DYAD with **Adam Bozson** and **Steven Hamblin**.
- Three invited talks.
  1. **Vicenzo Cutrona** (fromer PhD Student Università degli Studi di Milano - Bicocca)
  2. **Alexandra Shatova** (MSc Data Science @ City)
  3. **Valentina Carapella** (Data Scientist at Perspectum)
- Overview of graph database solutions.

Overview of Graph Database Solutions

Ernesto Jiménez-Ruiz
Lecturer in Artificial Intelligence

Introduction

Advanced graph database solutions:

- Scale to large Knowledge Graphs.
- Sophisticated indexing structures.
- Optimised reasoning.
- Fast query performance.
- Server solution in production.

State-of-the-art solutions

Dimensions:

- Free version.
- Compliance with Semantic Web standards.
- Reasoning capabilities.
- In-memory or In-disk.
- Documentation and installation requirements.
- Additional features.

A Survey of RDF Stores & SPARQL Engines for Querying Knowledge Graphs. arXiv:2102.13027 2021 (Appendix A)

Semantic Web standards

- The **World Wide Web Consortium (W3C)** is an international community that develops open standards to ensure the long-term growth of the Web: `https://www.w3.org/`
- On the Web and **beyond**.
- **Why standards?**
  - broader industry (and academic) agreement,
  - interoperability across organizations and applications,
  - avoids vendor lock-in of a particular (exchange or query) format.

Apache Jena TBD

- Free solution.
- Provides a native (in-disk) RDF store.
- In combination with Jena Fuseki to provide SPARQL Endpoint support.
- ✗ Supports reasoning as in Jena, but not direct support for OWL 2 nor the OWL 2 profiles.

`https://jena.apache.org/documentation/tdb/`

OpenLink Virtuoso

- ✓ Provides the SPARQL endpoint for DBpedia.
- – Open source and commercial versions.
- – Object-oriented database model.
- ✓ Native graph model storage providers for Jena and RDF4J.
- ✗ Custom inference rules. Partial support for OWL 2.

`https://virtuoso.openlinksw.com/`

`http://vos.openlinksw.com/owiki/wiki/VOS`

Blazegraph

- ✓ Provides the SPARQL Endpoint for Wikidata.
- ✓ Free and open source.
- – Both in-memory and disk-oriented storage.
- ✗ Only supports OWL 1 Lite reasoning.

`https://blazegraph.com/`

AllegroGraph

– Free and commercial licenses.

✓ Support for OWL 2 RL materialization.

✓ Client interface in several languages.

– Can be used to query both documents and graph data (via SPARQL).

`https://allegrograph.com/`

RDFox

– Commercial system. Free academic license on request.

✓ Support for materialization-based datalog reasoning (including OWL 2 RL and SWRL rules).

– In-memory RDF engine.

✓ Can be used via a Java API or remotely via a REST API or a SPARQL Endpoint.

✗ Limitation on the size of the memory.

`https://www.oxfordsemantic.tech/product`

Amazon Neptune

- Cloud-based only solution.
- Blazegraph is now part of Amazon Neptune.
- ✗ On-Demand pricing.
- ✗ Inferencing is not yet supported.

`https://aws.amazon.com/neptune/`

Neo4j

- ✓ Open source graph database.
- – Based on the Property Graph Model.
- ✗ Cypher as graph query language (no SPARQL support).
- – Support *via a plugin* for RDF, RDFS and OWL vocabularies.
- ✗ Basic inferencing support.
- ✓ Support for Analytics.
- ✓ Interfaces in many languages.

`https://neo4j.com/`

GRAKN.AI

✓ Grakn is an open-source, distributed knowledge graph database.

✓ Support for analytics.

✓ Interesting integration with machine learning models.

– Provides inferencing support.

✗ No support for any of the Semantic Web standards.

`https://grakn.ai/`
GRAKN.AI vs Semantic Web standards (some justifications that I do not personally share):
`https://blog.grakn.ai/knowledge-graph-representation-grakn-ai-or-owl-506065bd3f24`

GraphDB (formerly OWLIM)

- – Free and commercial versions.
- ✓ Very easy to install and use.
- ✓ Powerful reasoning features: including OWL 2 QL and RL profiles.
- – Includes text indexing via lucene.
- – Powered the early Linked Data services at the BBC.
- – **Our choice for the lab.**

`https://www.ontotext.com/products/graphdb/`

RDF4J (formerly Sesame)

✓ General Java framework to manage RDF data.

– Provides native (in-memory and in-disk) storage solutions.

– Can connect to other RDF stores *e.g.*,: Blazegraph, Amazon Neptune, GraphDB, Virtuoso.

– Indexing, reasoning and query processing techniques depend on the underlying storage engine.

`https://rdf4j.org/`

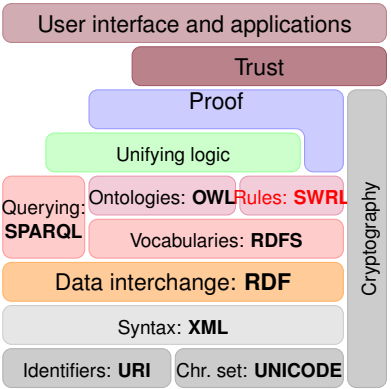
Additional material

SPARQL visual interfaces

https://sws.ifi.uio.no/project/optique-vqs/

Datalog and the Semantic Web Rule Language (SWRL)

$\texttt{hasParent(?x1,?x2)} \wedge \texttt{hasBrother(?x2,?x3)} \rightarrow \texttt{hasUncle(?x1,?x3)}$



Chapter 10 Pizza Tutorial: `https://www.michaeldebellis.com/post/new-protege-pizza-tutorial`

SHACL: Shapes Constraint Language

– A language to define constraints to validate RDF data.

✓ SHACL focus on CWA and provides a rich language to explicitly define checks over the data.

✗ Some of the constraints could be defined using OWL or rules (OWA), or SPARQL (CWA).

Chapter 11 Pizza Tutorial: `https://www.michaeldebellis.com/post/new-protege-pizza-tutorial`

RDF* and SPARQL*

- Uses the embedding triple operator « »
- Compact solution easier to read than reification
- Closer to Property Graphs

```
:ernesto :teaches :INM713 .
<<:ernesto :teaches :INM713>> dbpo:year "2021"^^xsd:gYear .
```

O. Hartig and B. Thompson. Foundations of an Alternative Approach to Reification in RDF. CoRR, abs/1406.3399, 2019
Additional Resources (March 2021):
`http://www.lotico.com/index.php/Metadata_for_RDF_Statements:_The_RDF-star_Approach`
`https://youtu.be/ZNfq12mdnsM`

Acknowledgements

- Today's invited speakers.
  - Vicenzo Cutrona
  - Alexandra Shatova
  - Valentina Carapella
- Ontotext/GraphDB.

Laboratory

Using the RDF store GraphDB

- Creating repositories.
- Loading data and ontology.
- Querying the data.