Knowledge Graph Publication and Browsing Using Neo4J

Ghislain Atemezing, Anh Huynh

@gatemezing
About Mondeca

• Software for taxonomy management, content analytics, knowledge discovery
• +20 years experience
• Agile, flat structure focused on software development
• Based in Paris / France
• A strong foothold in Europe and in North America
Agenda

- Motivation
- Our proposal
  - Mondeca Mappings RDF → PG
  - Implementation & Evaluation
  - KBrowser tool
  - Application Use cases
- Discussion
- Conclusions
Motivation: RDF vs Property Graph (PG)

• No properties on edge
• Formal semantics for inference
• Built on W3C standards process
• Multiple serialization format
• Schema languages (RDFS, OWL)
• Great for Linked (open) Data with no central control
• Great for Knowledge Graphs (KG) creation and publication

• Key-value pairs for nodes and edges
• No formal model representation
• Different vendors
• Cypher, PGQL, Gcore, GQL (standard to come)
• No schema language
• Great for graph analytics and path search
• Great for using ML libraries and graph traversal language (Gremlin)
Graph Database Market Update 2020 - 2021

• Neo4j is the leader for PG database  

• Ontotext GraphDB is the champion for RDF database

Bloor Research - https://www.bloorresearch.com/technology/graph-databases/#vendorlandscape
How to use a PG database (aka Neo4J) to publish and visualize Knowledge Graphs on the Web?
Our contributions

• Mondeca Mappings RDF→PG (Neo4J)
• Implementation & Evaluation
• KBrowser tool
• Application Use cases
Rule 1:
Subjects of triples are mapped to nodes in Neo4j. A node in Neo4j representing an RDF resource will be labeled :Resource and have a property uri with the resource’s URI.

\[(S, P, O) \Rightarrow (\text{:Resource}\{\text{value: S, kind:uri| bnode}\}) \land (\text{:Resource}\{\text{value: O, kind :uri| bnode}\})\]
Rule 2:
Predicates of triples are mapped to relationships in Neo4j if the object of the triple is a literal.

\[(S, P, O) \land \text{isLiteral}(O) \Rightarrow (\text{:Resource}{\text{value}: S, \text{kind:} \text{uri}}) -[:P] - (\text{:Resource}{\text{value}: O, \text{kind:} \text{literal}})\]
Rule 3:

Predicates of triples are mapped to relationships in Neo4J if the object of the triple is a resource.

\[(S, P, O)\&\&!\text{isLiteral}(O) \Rightarrow (:\text{Resource}\{\text{value: } S, \text{kind:uri}\}-[:P]->(:\text{Resource}\{\text{value: } O, \text{kind:uri}\}))\]
Mappings sample visual result
Experiment: Loading time performance

- Neo4j 3.5.6 community edition memory heap size of 8GB and page cache memory of 1GB
- Creation of 5 indexes p, c, cp, kind and value
- Call the procedure with three parameters: the file, the serialization, and the number of triples per transaction set to 50K triples

- Datasets: one vocabulary and random datasets from DBpedia dumps from 2015 [1]

<table>
<thead>
<tr>
<th>Dataset</th>
<th>NbTriples</th>
</tr>
</thead>
<tbody>
<tr>
<td>schema.org</td>
<td>15,400</td>
</tr>
<tr>
<td>DBpedia citations links EN</td>
<td>880,250</td>
</tr>
<tr>
<td>DBpedia category labels EN</td>
<td>1,321,139</td>
</tr>
<tr>
<td>DBpedia instance types EN</td>
<td>5,067,876</td>
</tr>
<tr>
<td>DBpedia person data EN</td>
<td>8,422,913</td>
</tr>
<tr>
<td>DBpedia out degree EN</td>
<td>11,998,171</td>
</tr>
</tbody>
</table>

KBROWSER Architecture overview

KBrowser UI

KBrowser REST API
search, repository, graph, etc

KBrowser engine

KB - Technology Stack

UI
GWT (JS), DJ, bootstrap CSS responsive

API
REST API

Web Server
Tomcat 9

Graph Database
Neo4J (Cypher as internal query language)

Search Engine
Elasticsearch

Document Database
MongoDB

KB engine
Spring framework (Springboot, Spring MVC)

Exchange Format
RDF

Installation
Ansible Scripts: Installation, Backup, Restore, Version upgrade, Uninstallation
Docker container image

RDF files
From ITM or other sources

SCG Workshop 2021 @SEMANTiCS- Amsterdam - September 6th, 2021
KBrowser Applications: TerMef and French BioloInc Portal

• Live deployments of KBrowser
• Easy publication of terminologies for both domains (Finance and Health)
• Search interface and graph views of relationships (Molecular visualization)
• KBrowser runs on a minimal RAM (2GB)
Limitations

• Benchmark our proposal with existing solutions (NSMNTX, Angles et. al, 2020)
• Check memory evolution with billions of triples
• Integrate named graphs in the proposal
• Target other graphs databases (ArangoDB, JanusGraph)
• We don’t provide mappings from PG → RDF data models
Conclusions

• We have proposed a generic mappings from RDF2PG data model for Neo4J
• We have implemented the procedure to load RDF datasets using Neo4J
• The loading time is promising in an 8G RAM machine
• We use the mappings to showcase KBrowser, a tool to publish and visualize KGs using Neo4J
• KBrowser is featured in two portals showcasing the validity and usability of our approach
Thank you for your attention!
Knowledge Graph Publication and Browsing Using Neo4J

Ghislain Atemezing, Anh Huynh

@gatemezing