

Knowledge Graph Publication and Browsing Using Neo4J

Ghislain Atemezing, Anh Huynh

 @gatemezing

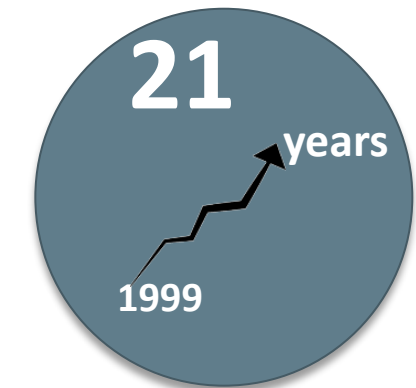


M O N D E C A



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





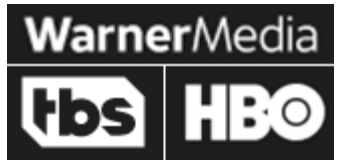
Key references



International Olympic Committee



American Association of Family Physicians



Uppsala University Hospital





Agenda

- Motivation
- Our proposal
 - Mondeca Mappings RDF → PG
 - Implementation & Evaluation
 - KBrowser tool
 - Application Use cases
- Discussion
- Conclusions



Motivation: RDF vs Property Graph (PG)

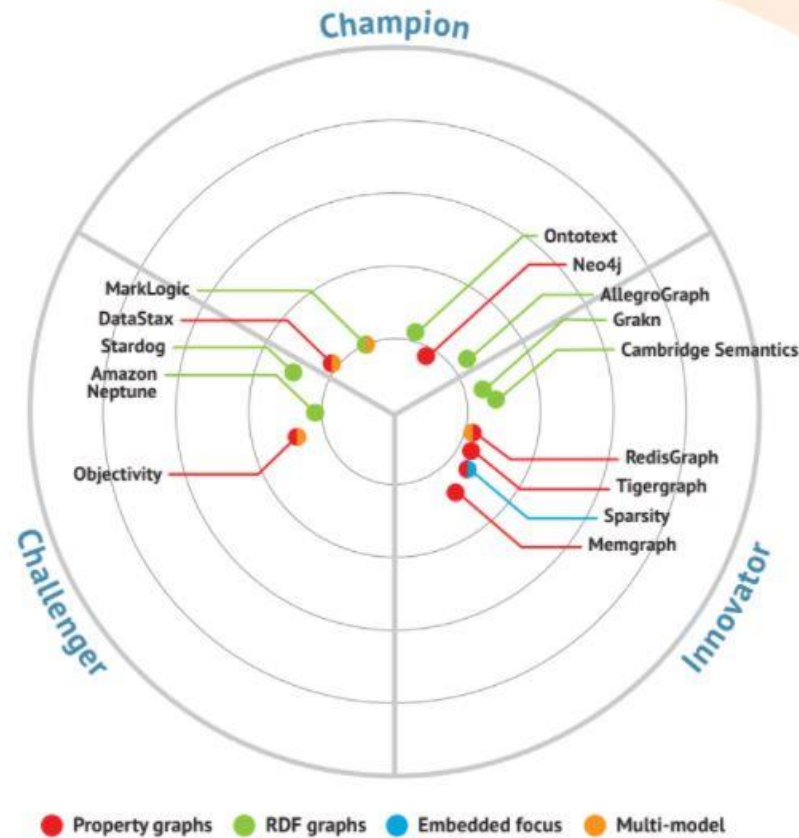
- No properties on edge
- Formal semantics for inference
- Built on W3C standards process
- Multiple serialization format
- Schema languages (RDFS, OWL)
- Great for Linked (open) Data with no central control
- Great for Knowledge Graphs (KG) creation and publication
- Key-value pairs for nodes and edges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



Our main business use case

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



Mappings RDF → PG: Node creation

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.

$(S, P, O) \Rightarrow (:Resource\{value: S, kind:uri | bnode\})$

&&

$(:Resource\{value:O, kind :uri | bnode\})$

```
dbr:Mondeca rdf:type  
dbo:Organisation.
```



Resource
value="dbr:Mondeca" kind="uri"

Resource
value="dbo:Organisation" kind="uri"



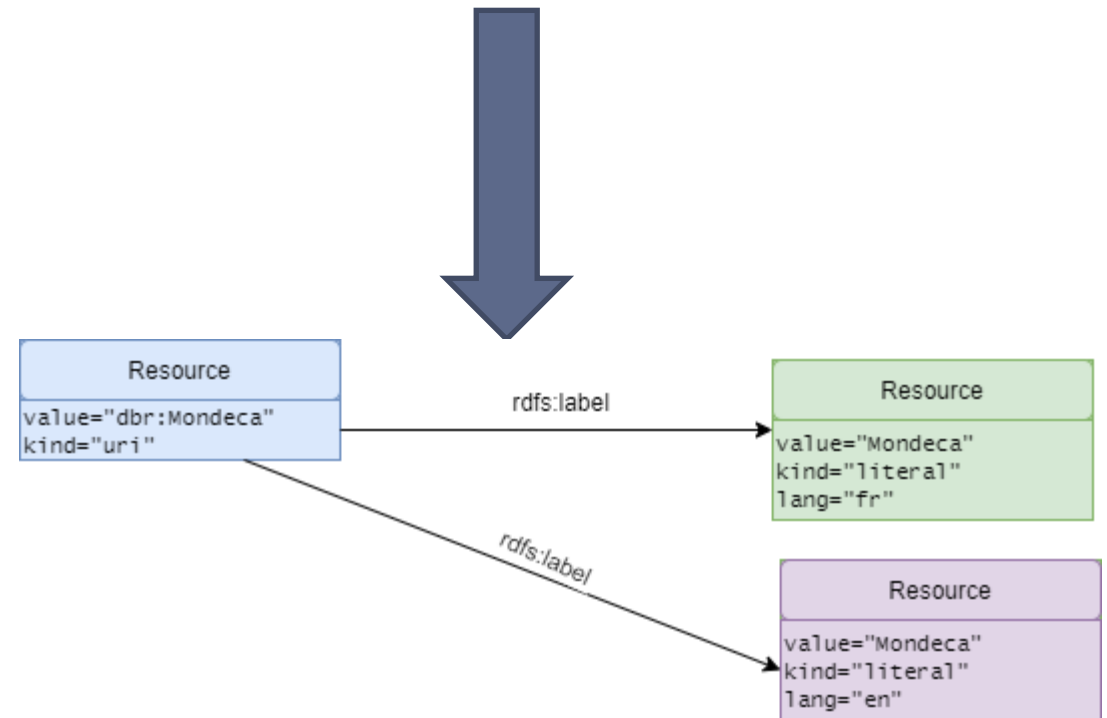
Mappings RDF → PG: Datatypes and multilingual information

Rule 2:

Predicates of triples are mapped to relationships in Neo4j if the object of the triple is a literal.

$(S, P, O) \ \&\& \text{isLiteral}(O) \Rightarrow (:$
 $\text{Resource}\{\text{value}:S, \text{kind}:\text{uri}\})$
 $-[:P] \rightarrow (:\text{Resource}\{\text{value}:O,$
 $\text{kind}:\text{literal}\})$

```
dbr:Mondeca rdfs:label  
"Mondeca"@fr , "Mondeca"@en.
```





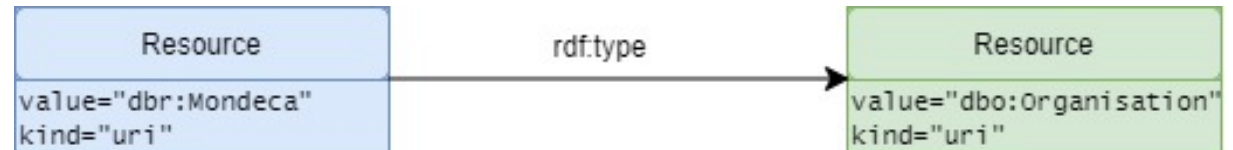
Mappings RDF → PG

Rule 3:

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

```
(S, P, O) && !isLiteral(O) => (:  
Resource{value: S, kind:uri})-[:  
P]->(:Resource{value: O, kind:  
:uri})
```

```
dbr:Mondeca rdf:type  
dbo:Organisation.
```





Mappings sample visual result



INDEXED <id>: 0 kind: uri value: http://dbpedia.org/resource/Mondeca

Resource

<http://dbpedia.org/ontology/Company>

<http://schema.org/Organization>

<id>: 0

http://dbpedia.org/ontology/abstract:
 Mondeca is a leading provider of software based on semantic technologies. It has been recognized by Gartner as a top vendor in the Semantic Web as well as in the Taxonomy and Ontology Management categories.

http://www.w3.org/2000/01/rdf-schema#label: Mondeca

uri: http://dbpedia.org/resource/Mondeca

http://dbpedia.org/ontology/foundingYear: 2000



Experiment: Loading time performance

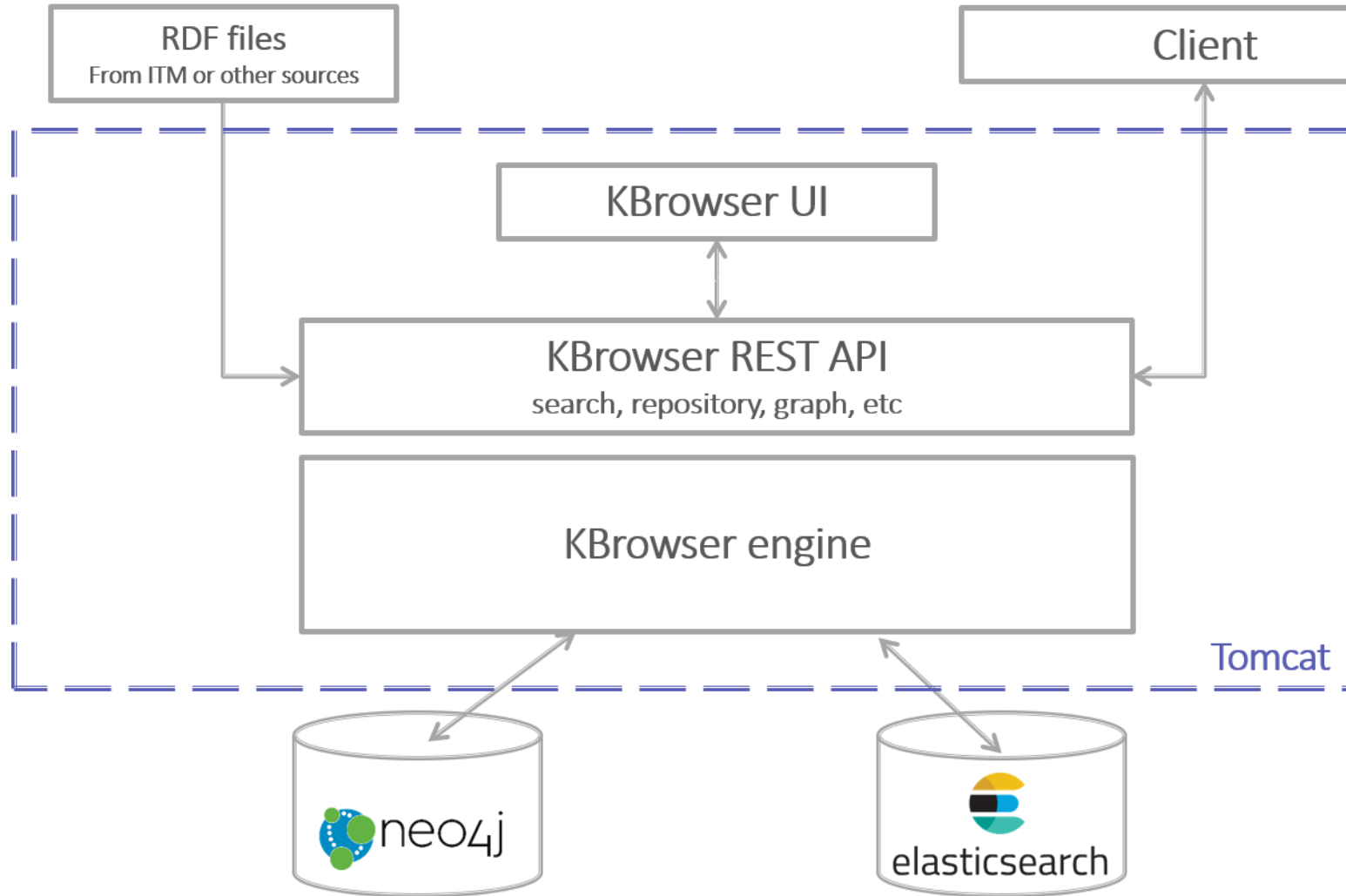
- Neo4J 3.5.6 community edition memory heap size of 8GB and page cache memory of 1GB
- Creation of 5 indexes p, c, cp, kind and value
- Call the procedure with three parameters: the file, the serialization, and the number of triples per transaction set to 50K triples
- Datasets: one vocabulary and random datasets from DBpedia dumps from 2015 [1]

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

[1] <http://downloads.dbpedia.org/2015-10/core-i18n/en/>



KBrowser Architecture overview



KB - Technology Stack

UI	GWT (JS), D3js, 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



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)

The screenshot shows the TerMef website in a browser window. The URL is terminologie.finances.gouv.fr/index. The page features a blue header with a search bar and a navigation menu. Below the header, there is a section titled "Accéder aux référentiels" (Access to references) which displays eight categories of terminologies, each with an icon, a title, and the number of elements:

Category	Number of Elements	Last Update
FINANCES PUBLIQUES	579	03/02/2021
AUTOMOBILE	150	04/08/2021
CHIMIE ET MATÉRIAUX	647	04/08/2021
ECONOMIE	1211	04/08/2021
NUCLÉAIRE	455	04/08/2021
ÉNERGIES FOSSILES	283	04/08/2021
INFORMATIQUE	686	04/08/2021
TÉLÉCOMMUNICATIONS	428	04/08/2021

At the bottom of the page, there is a footer with links for "À propos de TerMef", "Contact", "Mentions Légales", "Accessibilité", and "Plan du site". The browser's taskbar at the bottom shows the Windows logo, a search bar, and the system tray with the date 02/09/2021 and time 15:51.



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!

Q/A session



M O N D E C A

Knowledge Graph Publication and Browsing Using Neo4J

Ghislain Atemezing, Anh Huynh

 @gatemezing



M O N D E C A