Author(s): Rafael Barros Date: November 19th 2024

Welcome

RDF and Property Graphs Interoperability: Status and Issues

Authors: **Renzo Angles, Harsh Thakkar, Dominik Tomaszuk**

Affiliations: **Universidad de Talca, University of Bonn, University of Bialystok**

Introduction:

RDF and Property Graphs:

- Both are graph-oriented database models.
- RDF uses triples (subject, predicate, object).
- Property Graphs use nodes and edges with properties.

Objective: - Study interoperability between RDF and Property Graph databases.

Types of Interoperability

- **Syntactic Interoperability:**
  - Data exchange at the level of serialization formats.
- **Semantic Interoperability:**
  - Common understanding of data meanings.
- **Query Interoperability:**
  - Transforming queries between different query languages.

RDF Databases

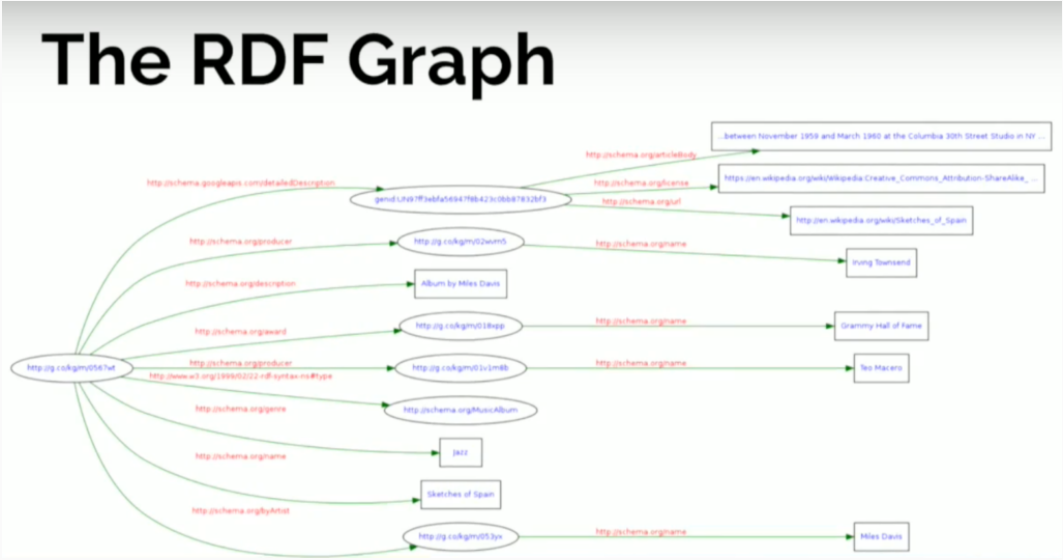
- **Data Model:**
  - RDF triples (subject, predicate, object).
  - Visualized as graphs.
- **Schema:**
  - RDF Schema, OWL, SHACL, ShEx.
- **Query Language:**
  - SPARQL.

RDF Data (Turtle syntax)

```
@prefix schema: <http://schema.org/> .
@prefix ns0: <http://schema.googleapis.com/> .

INSERT DATA {
        <http://g.co/kg/m/0567wt>
         schema:name "Sketches of Spain" ;
         a schema:MusicAlbum ;
         schema:description "Album by Miles Davis" ;
         schema:genre "Jazz" ;
         ns0:detailedDescription [
                 schema:license "Creative_Commons_Attribution-ShareAlike_3.0_License" ;
                 schema:url "http://en.wikipedia.org/wiki/Sketches_of_Spain" ;
                 schema:articleBody "...between Nov 1959 and Mar 1960 at the Columbia 30th St Studio in NY City" ] ;
         schema:award <http://g.co/kg/m/018xpp> ;
         schema:byArtist <http://g.co/kg/m/053yx> ;
         schema:producer <http://g.co/kg/m/01v1m8b>, <http://g.co/kg/m/02wvrn5> .
        <http://g.co/kg/m/018xpp> schema:name "Grammy Hall of Fame" .
        <http://g.co/kg/m/053yx> schema:name "Miles Davis" .
        <http://g.co/kg/m/01v1m8b> schema:name "Teo Macero" .
        <http://g.co/kg/m/02wvrn5> schema:name "Irving Townsend" .  }
```

- You can see that the triples are identified by a URI, which is the **subject**. The **predicate** is the name and the **object** will be *Sketches of Spain*, which together is a sequence of triples.

Let's look at how this information is displayed graphically:



Property Graph Databases

- **Data Model:**
  - Directed labeled multi-graph with properties.
- **Schema:**
  - Node types, edge types, integrity constraints.
- **Query Languages:**
  - No standard; examples include Cypher, PGQL, G-CORE.

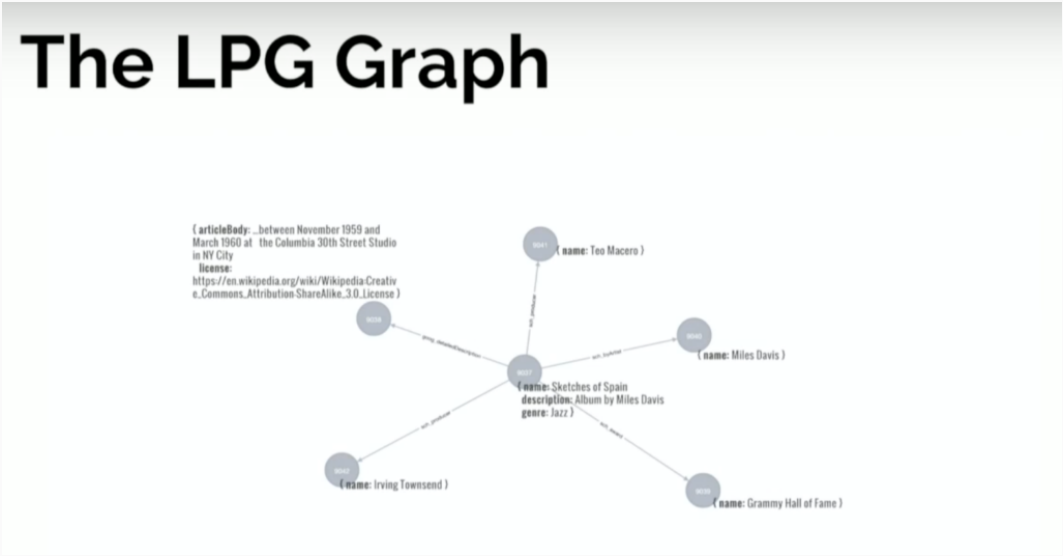
LPG Data (Cypher)

```
CREATE (sos:Resource:MusicAlbum { name: "Sketches of Spain",
                                  description: "Album by Miles Davis",
                                  genre: "Jazz"})

CREATE (dd { license: "Creative_Commons_Attribution-ShareAlike_3.0_License",
             url: "http://en.wikipedia.org/wiki/Sketches_of_Spain",
             articleBody: "...between Nov 1959 and Mar 1960 at the Columbia 30th St Studio in NY City"})

CREATE (sos)-[:goog_detailedDescription]->(dd)

CREATE (sos)-[:award]-> ({ name: "Grammy Hall of Fame" })
CREATE (sos)-[:byArtist]->({ name: "Miles Davis" })
CREATE (sos)-[:producer]->({ name: "Teo Macero" })
CREATE (sos)-[:producer]->({ name: "Irving Townsend" })
```

Figure 3: Cypher example

- The **semantics** are the same. There's no standard serialization format or a way of expressing a labeled property graph, but rather a sequence of `CREATE` statements do the job here.

Let's look at how this information is displayed graphically:



- **Core Difference:** nodes have this internal structure and values of attributes don't represent vertices in the graph.

Syntactic Interoperability

- **Challenges:**
  - No standard data format for Property Graphs.
  - Different serialization formats (e.g., Turtle, RDF/XML for RDF).
- **Approaches:**
  - Textual mappings, intermediate data formats.

Semantic Interoperability

- **Challenges:**
  - Special RDF features (e.g., blank nodes, reification).
  - Partial schemas in RDF.
- **Approaches:**
  - Data and schema transformation methods.
  - Use of transformation languages (e.g., XSPARQL, RML).

Query Interoperability

- **Challenges:**
  - Lack of a standard query language for Property Graphs.
  - Different paradigms (declarative vs. imperative).
- **Approaches:**
  - Tools like Gremlinator for SPARQL to Gremlin translation.
  - Efforts to standardize Property Graph query languages.

Issues and Challenges

- **Syntactic Interoperability:**
  - No standard data format for Property Graphs.
  - Differences in serialization formats.
- **Semantic Interoperability:**
  - RDF features not easily modeled in Property Graphs.
  - Need for schema discovery and transformation.
- **Query Interoperability:**
  - No standard query language for Property Graphs.
  - Different query paradigms and semantics.

Conclusions

- **Importance of Interoperability:**
  - Facilitates data exchange, integration, and reuse.
- **Future Directions:**
  - Standardization of Property Graph data model and query language.
  - Development of robust transformation methods.

Acknowledgments

- **Funding:**
  - Millennium Institute for Foundational Research on Data.
  - EU H2020 R&I project BOOST4.0.
  - National Science Center, Poland.

References:

- About 48 written in the Article

Thank you

- **Questions?**