Geospatial Querying in Apache Marmotta

Sergio Fernández (Redlink GmbH)
ApacheCon Big Data
Vancouver, May 11th 2016
Sergio Fernández

@wikier
https://linkedin.com/in/sergiofernandez
https://www.wikier.org

Partner Technology Manager at Redlink GmbH

also...
External Lecturer at Fachhochschule Salzburg
Member of The Apache Software Foundation
Geospatial Querying in Apache Marmotta™
What is Apache Marmotta?

- **An Open Platform for Linked Data**, an open implementation of a Linked Data Platform that can be easily used, extended and deployed by organizations who want to publish Linked Data or build custom applications on Linked Data.

- **Key features:**
  - Read-Write Linked Data server
  - RDF triple store with transactions, versioning and rule-base reasoning
  - LDP, SPARQL and LDPath querying
  - Transparent Linked Data caching
What is Linked Data?

- The Semantic Web is a **Web of Data**

- Semantic Web technologies (**RDF**, OWL, SKOS, SPARQL, etc.) provide an environment where applications can query that data, draw inferences using vocabularies, etc.

- **Linked Data** lies at the heart of what **Semantic Web** is all about: large scale integration of, and reasoning on, data on the Web.

- A typical case of a large Linked Dataset is **DBPedia**, which, essentially, makes the content of Wikipedia available as Linked Data.
What is RDF?

- The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model.

- RDF is **directed labeled graph**, where:
  - nodes are resources;
  - edges represent the named links between two resources;
  - the composition of one resource (subject) linked (with a predicate) to another (object) is known as "RDF triple";
  - a set of triples form a RDF graph.
Currently Marmotta provides three main means of querying:

- **LDP 1.0 (Linked Data Platform),**
  - a W3C protocol based on HTTP for managing Linked Data resources
  - [https://www.w3.org/TR/ldp/](https://www.w3.org/TR/ldp/)

- **SPARQL 1.1 (SPARQL Protocol and RDF Query Language)**
  - a W3C RDF query language and protocol
  - [https://www.w3.org/TR/sparql11-query/](https://www.w3.org/TR/sparql11-query/)

- **LDPath**
  - a path language for Linked Data, similar to XPath for XML
  - [https://marmotta.apache.org/ldpath/language](https://marmotta.apache.org/ldpath/language)
GeoSPARQL

- The **OGC GeoSPARQL** standard supports representing and querying geospatial data on the Semantic Web.
- GeoSPARQL defines a vocabulary for representing geospatial data in RDF, and a SPARQL extension for processing geospatial data.
- It makes use of both WKT (Well Known Text) and GML for representing geometries as literals.
There are three key classes in the GeoSPARQL ontology:

- geo:SpatialObject
  - a superclass of both Features and Geometries;

- geo:Feature
  - a thing that can have a spatial location; e.g., a park;

- geo:Geometry
  - a representation of a spatial location; i.e., a set of coordinates.

Namespace: http://www.opengis.net/ont/geosparql#
GeoSPARQL basic data model
1. **Materialization**
   - Pros: fast querying
   - Cons: materialization is computationally expensive, requires more storage capacity and native operators

2. **Query translation**
   - Pros: direct comparison, optimal storage and no need of native operators
   - Cons: slow querying

In Marmotta we decided to go for the first option.
GeoSPARQL in Marmotta

- More precisely we should say "GeoSPARQL in KiWi"
  - KiWi is our triple store based on RDBMS
  - Marmotta also supports many other Sesame-based triple stores as backend

- Support implemented based on PostGIS for PostgreSQL
  - Support not available for other databases

- All further technical details available at the wiki: https://wiki.apache.org/marmotta/GSoC/2015/MARMOTTA-584

- Documentation at http://marmotta.apache.org/kiwi/geosparql
GeoSPARQL example

Simple query to get all geometries that are contained by other.

Particularly this example queries for the first ten municipalities in the region of Madrid.
Another query to get all geometries that are touch other.

Particularly this example queries for the rivers that make borderline with Austria.
Another query to get all geometries that are crosses another.

Particularly this example queries for all Mountain bike routes that cross a city.
GeoSPARQL coverage

Apache Marmotta 3.4.0 will(*) support:

- Simple Features Topological Relations
- Egenhofer Topological Relations
- RCC8 Topological Relations
- Non-Topological Functions

Better to check the documentation: https://wiki.apache.org/marmotta/GSoC/2015/MARMOTTA-584

(*) still under development at MARMOTTA-584 branch, to be released in the upcoming weeks
let's demo!
Thanks!
we’re HIRING!

http://redlink.co/company/careers
The work presented here has been developed in the context of the TourPack project, partially funded the Austrian Research Promotion Agency (FFG) IKT der Zukunft program under grant agreement no. 845600.

Acknowledgements

Thanks to the student Francisco Xavier Sumba Toral for contributing the initial GeoSPARQL implementation as part of his project during the Google Summer of Code 2015.

Thanks to Google for such awesome open source program!