Ontop-spatial: Geospatial Data Integration using GeoSPARQL-to-SQL Translation

Konstantina Bereta[1], Guohui Xiao[2], Manolis Koubarakis[1], Martina Hodrius[3], Conrad Bielski[4], and Gunter Zeug[5]

[1] National and Kapodistrian University of Athens, Greece
[2] Free University of Bozen-Bolzano, Italy
[3] VISTA GmbH, Germany
[4] EOXPLORE UG, Germany
[5] Terranea UG, Germany

Abstract. We present *Ontop*-spatial, a geospatial extension of the well-known OBDA system Ontop, that leverages the technologies of geospatial databases and enables GeoSPARQL-to-SQL translation. We showcase the functionalities of the system in real-world use cases which require data integration of different geospatial sources.

1 Introduction

During the recent years, the amount of geospatial data in the Web of Data has increased. This is because geospatial data practitioners coming from various domains (e.g., earth scientists, geologists, civil engineers) that are involved in the processing of geospatial data, also publish them as RDF to increase its value by combining it with other data.

As a result, the Semantic Web community became active proposing data models, query languages and applications for the representation, modeling and visualization of linked geospatial data [5]. These efforts have been strengthened by the establishment of the Open Geospatial Consortium (OGC) standard GeoSPARQL [1], a geospatial extension of RDF and SPARQL. At the same time, similar extensions of RDF and SPARQL were also proposed, such as the framework of stRDF and stSPARQL which extends RDF and SPARQL with space and time [6,3]. RDF stores with geospatial support were also implemented, such as Parliament, uSeekM and Virtuoso, that implement a subset of GeoSPARQL, and Strabon [6], that implements both GeoSPARQL and stSPARQL. Despite the long tradition of research in geospatial relational databases, and the existence of Ontology-Based Data Access (OBDA) systems that offer on-the-fly SPARQL-to-SQL translation based on ontologies and mappings (e.g., Ontop [9], Ultrawrap [10], Morph [8]), there was no OBDA system with GeoSPARQL support. In [2], we describe how we extended Ontop [9] with geospatial support and implemented *Ontop*-spatial, a geospatial extension of the system Ontop.

The development of *Ontop*-spatial was initially motivated by the Statoil use case in the context of the EU FP7 project OPTIQUE[6], in order to address the

---

[6] http://optique-project.eu/

issue of creating virtual RDF graphs on top of large relational databases that contain geometries and get frequently updated. *Ontop*-spatial is being used in the Urban accountant, Land management, and Crisis Mapping services of the EU FP7 project MELODIES[7]. More recently, *Ontop*-spatial has been used in the maritime security domain by the German BMBF project EMSEC [4].

In this demo paper we focus on how *Ontop*-spatial can be used to integrate geospatial data from different sources and express rich queries that combine them. In Section 2 we present a technical overview of *Ontop*-spatial, explaining its compliance with other visualization tools that are useful for domain experts. In Section 3 we give an overview of our demonstration, presenting how *Ontop*-spatial is used in land management and crisis mapping real-world scenarios.

2   *Ontop*-spatial

Ontop-spatial[8] extends Ontop to enable the on-the-fly GeoSPARQL-to-SQL translation on top of geospatial databases and thus becomes the first OBDA system with geospatial support. It is able to connect to a geospatial database (currently PostGIS or Spatialite) and create virtual geospatial RDF graphs on top of it, using ontologies and mappings. It supports the following components of GeoSPARQL: *Core*, *Topology Vocabulary*, *Geometry topology extension*, *RDFS entailment* and a subset of *Geometry Extension*. To the best of our knowledge, it is also the first GeoSPARQL implementation that supports the *query rewrite extension* of GeoSPARQL. In [2] we explain how GeoSPARQL queries are processed by *Ontop*-spatial and are transformed into the respective spatial SQL queries that are evaluated by geospatial databases.

For example, the GeoSPARQL query in Listing 1.1 retrieves buildings that are affected by floods (i.e., they have intersecting geometries with the flood geometries). The query described in Listing 1.2 can also be evaluated in *Ontop*-spatial returning the same results, as it gets transformed internally to the query in 1.1, according to the *query rewrite extension* of the GeoSPARQL specification.

Listing 1.1: Query 1 (Quantitative)

```
SELECT DISTINCT ?name ?build ?type
WHERE {?s1 f:type ?type .
 ?s1 geo:asWKT ?g1 .
 ?s2 geo:asWKT ?g2 .
 ?s2 rdf:type osm:Building .
 ?s2 osm:hasName ?name .
 ?s2 osm:buildingCategory ?build .
FILTER(geof:sfIntersects(?g1, ?g2)) }
```

Listing 1.2: Query 2 (Qualitative)

```
SELECT DISTINCT ?name ?build ?type
WHERE {
 ?s1 f:type ?type .
 ?s2 rdf:type osm:Building .
 ?s2 osm:hasName ?name .
 ?s2 osm:buildingCategory ?build .
 ?s1 geo:sfIntersects ?s2
}
```

---

[7] http://www.melodiesproject.eu/software-tools
[8] https://github.com/ConstantB/ontop-spatial

3 Demonstration Overview

The demonstration of *Ontop*-spatial will be based on the real-world scenarios of land management and crisis mapping studied in the EU FP7 project MELODIES. These use cases are led by the German company VISTA and the companies EOXPLORE and Terranea respectively.

In the land management scenario, we are interested in discovering agricultural fields that intersect with protected areas. To achieve this, we need to combine information from the following geospatial datasets:

- Agricultural fields. This dataset contains information about fields, i.e., their geographic location, their name, code, etc. This is in-house data, but we can use a sample of it for the demonstration.
- Protected areas. This is a set of different datasets that describe categories of protected areas. Most of this data is in-house.
- Corine Land Cover (CLC). This dataset is released as open data by the European Environment Agency (EEA) and it contains information about the land cover of various countries of Europe. In the context of this use case, we focus on categories that represent protected areas.

In the crisis mapping scenario, we are mainly interested in information about floods, which is stored originally in a PostGIS database[9] maintained by EOXPLORE and Terranea. This database consists of a table that contains only basic information about floods such as the location, date and some information about the region and the country. We enrich this dataset by integrating the following open datasets:

- The Global Administrative Areas dataset[10] that contains information about all levels of administrative divisions worldwide.
- The Corine Land Cover dataset, focusing in areas that are characterized as "Water bodies".
- The Open Street Maps dataset (OSM). OSM also contains some categories for water bodies, such as rivers and lakes, as well as points of interest.

All datasets described in both use cases are originally in Shapefile format, except for the floods data that are relational (stored in a PostGIS database).

In the demonstration we will show how we can integrate geospatial data coming from different sources and then pose rich queries combining them, in both of these use cases.

Integration. All Shapefiles are imported to a database. In case a database pre-exists (e.g., the floods database), so we import the additional shapefiles there. Every shapefile will be imported to a corresponding table, and the columns of this table will be the same as the attributes of the shapefile.

Ontology. We need to construct an ontology to model the information that we want to map to RDF. In order to exploit the geospatial features, this ontology should be an extension of the GeoSPARQL ontology[11].

---

[9] `http://bit.ly/29otBQk`

[10] `http://www.gadm.org/`

[11] `http://schemas.opengis.net/geosparql/1.0/geosparql_vocab_all.rdf`

Mappings. Mappings describe how relational data can be translated to RDF. A mapping file is constructed, using either the R2RML mapping languages or the native mapping language of *Ontop*. The Protègè plugin of Ontop provides a user-friendly graphical interface for editing and managing mappings.

Posing rich geospatial queries. A user can create a repository using either the embedded Sesame-based web interface of Ontop or the visual query interface the Optique platform [11].

Visualization of results. *Ontop*-spatial inherits the ability of *Ontop* to be used as a SPARQL endpoint. So the results can be visualized on the map using Sextant [7], a web-based tool for browsing and visualizing linked geospatial data, that is able to connect to (Geo)SPARQL endpoints and project the results of the geospatial queries on the map.

The queries described in 1.1 and 1.2 are examples of the queries used in the crisis mapping scenario. A more detailed description of the demonstration is given in the online video: `https://youtu.be/F5_2Zxi5_e8`.

References

1. Open Geospatial Consortium. OGC GeoSPARQL - A geographic query language for RDF data. OGC Candidate Implementation Standard (02 2012)
2. Bereta, K., Koubarakis, M.: Ontop of Geospatial Databases. In: International Semantic Web Conference (ISWC) 2016. (To appear).
3. Bereta, K., Smeros, P., Koubarakis, M.: Representation and querying of valid time of triples in linked geospatial data. In: ESWC. LNCS, vol. 7882, pp. 259–274 (2013)
4. Bruggemann, S., Bereta, K., Xiao, G., Koubarakis, M.: Ontology-Based Data Access for Maritime Security, pp. 741–757. Springer International Publishing (2016)
5. Koubarakis, M., Karpathiotakis, M., Kyzirakos, K., Nikolaou, C., Sioutis, M.: Data Models and Query Languages for Linked Geospatial Data. LNCS, vol. 7487, pp. 290–328. Springer (2012)
6. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: A Semantic Geospatial DBMS. In: ISWC. LNCS, vol. 7649, pp. 295–311. Springer (2012)
7. Nikolaou, C., Dogani, K., Bereta, K., Garbis, G., Karpathiotakis, M., Kyzirakos, K., Koubarakis, M.: Sextant: Visualizing time-evolving linked geospatial data. J. Web Sem. 35, 35–52 (2015)
8. Priyatna, F., Corcho, O., Sequeda, J.: Formalisation and experiences of R2RML-based SPARQL to SQL query translation using Morph. In: Proc. of the 23rd International Conference on World Wide Web. pp. 479–490. ACM, NY, USA (2014)
9. Rodriguez-Muro, M., Rezk, M.: Efficient SPARQL-to-SQL with R2RML mappings. Journal of Web Semantics (2015)
10. Sequeda, J., Miranker, D.P.: Ultrawrap: SPARQL execution on relational data. Web Semantics: Science, Services and Agents on the World Wide Web 22(0) (2013)
11. Soylu, A., Giese, M., Jiménez-Ruiz, E., Vega-Gorgojo, G., Horrocks, I.: Experiencing OptiqueVQS: a multi-paradigm and ontology-based visual query system for end users. Universal Access in the Information Society 15(1), 129–152 (2016)