

The Spatiotemporal RDF Store Strabon

Kostis Kyzirakos, Manos Karpathiotakis, Konstantina Bereta, George Garbis,
Charalampos Nikolaou, Panayiotis Smeros, Stella Giannakopoulou, Kallirroï
Dogani, and Manolis Koubarakis

National and Kapodistrian University of Athens, Greece
{kkyzir,mk,Konstantina.Bereta,ggarbis}@di.uoa.gr
{charnik,psmeros,sgian,kallirroï,koubarak}@di.uoa.gr

Abstract. Strabon is a very scalable and efficient RDF store for storing and querying geospatial data that changes over time. We present the geospatial and temporal features of Strabon and we demonstrate their utilization in the fire monitoring and the burn scar mapping applications of the National Observatory of Athens.

1 Introduction

In this paper we present the spatiotemporal features of the system Strabon¹ and their utilization in the fire monitoring and the burn scar mapping applications of the National Observatory of Athens (NOA), which have been developed in the context of the project TELEIOS².

Strabon is a semantic geospatial DBMS for storing and querying geospatial data that changes over time. It implements the data model stRDF, the query language stSPARQL and the respective part of the OGC standard GeoSPARQL [1]. The data model stRDF is an extension of the W3C standard RDF for representing time-varying geospatial data. The query language stSPARQL is an extension of the query language SPARQL 1.1 and it has been implemented in Strabon offering scalability to billions of stRDF triples. The initial versions of the data model stRDF and the query language stSPARQL have been described in [5]. The geospatial features of the most recent versions of stRDF and stSPARQL, and their implementation in Strabon have been presented in [7]. This work also presents the architecture of the system, the optimizations followed for storing and querying geospatial data, and an experimental evaluation which shows that Strabon in most cases performs better than any other geospatial DBMS that has been competed with. The valid time dimension of the data model stRDF and the query language stSPARQL has been described in detail in [2].

We were motivated by the fire monitoring application of NOA to extend RDF and SPARQL 1.1 with the ability to store and query geospatial information that changes over time. In the fire monitoring service stRDF is used to represent satellite image metadata (e.g., time of acquisition, geographical coverage), knowledge extracted from satellite images (e.g., a certain image region is a

¹ <http://strabon.di.uoa.gr/>

² <http://www.earthobservatory.eu/>

hotspot) and auxiliary geospatial datasets encoded as linked data. The hotspot products are encoded to stRDF, so that they can be combined with auxiliary linked geospatial data. In this application, the user-defined time dimension is used as the *detection time* of hotspots.

The burn scar mapping (BSM) application of NOAA motivated us to design the valid time dimension of stRDF and stSPARQL and extend Strabon with valid time support, in order to be able to retrieve the evolution of the land cover of areas through time. Strabon is one of the very few RDF stores that supports storing and querying the valid time of triples.

This document is structured as follows. In Section 2 we describe the functionalities and the architecture of the system Strabon. In Section 3 we present the use of Strabon in the fire monitoring application of NOAA and in Section 4 we present the use of Strabon in the burn scar mapping application of NOAA. In Section 5 we describe the scenarios that will be used to demonstrate the spatiotemporal functionalities of the system.

2 The Spatiotemporal Features of Strabon

The system Strabon implements the most recent versions of stRDF and stSPARQL. In the new version of stRDF, we use the widely adopted OGC standards Well Known Text (WKT) and Geography Markup Language (GML) to represent geospatial data as literals of datatype `strdf:geometry`³. The new version of stSPARQL extends SPARQL 1.1 with the machinery of the OGC-SFA standard. We achieve this by defining one URI for each of the SQL functions defined in the standard and use them in SPARQL queries. Similarly, we have defined a Boolean SPARQL extension function for each topological relation defined in OGC-SFA (topological relations for simple features), [4] (Egenhofer relations) and [3] (RCC-8 relations). In this way stSPARQL supports multiple families of topological relations our users might be familiar with. Using these functions stSPARQL can express topological relations between geometry objects in the select or in the filter part of the query.

To be able to represent periods, we have also introduced the `strdf:period` datatype. The valid time annotation of a triple can be added at the end of the triple and it can be either a literal of the `strdf:period` or the `xsd:dateTime` datatypes. The temporal constants `NOW` and `UC` (i.e., “Until Changed”) are also introduced. The first denotes the current timestamp and the second the time persistence of a triple when placed as the ending time of a period. The query language stSPARQL also defines a wide variety of temporal functions to express relations between intervals, or instants and intervals. Period constructors and temporal aggregates are also defined. A full reference of the spatial and temporal functions provided in stSPARQL can be found online⁴.

Strabon 3.0 is a fully-implemented, open-source, storage and query evaluation system for stRDF/stSPARQL and the corresponding subset of GeoSPARQL.

³ <http://strdf.di.uoa.gr/ontology>

⁴ <http://www.strabon.di.uoa.gr/stSPARQL>

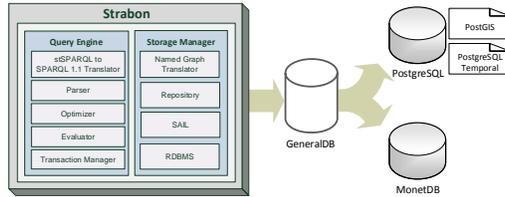


Fig. 1. Architecture of the system Strabon

Strabon has been implemented by extending the well-known RDF store Sesame transparently so that it can be compatible with most recent versions of Sesame. It also uses a DBMS as a backend, which can be a spatially and temporally enabled PostgreSQL database. MonetDB can also serve as backend for the geospatial features only. As Figure 1 shows, Strabon also consists of the *storage manager* and the *query engine*.

The storage manager stores stRDF triples using the per predicate scheme of Sesame and dictionary encoding. For each predicate table, two B+ tree two-column indices are created. For each dictionary table a B+ tree index on the id column is created. The geometries of spatial literals are stored as values of the `geometry` datatype in a separate table that uses an R-tree-over-GiST spatial index on the respective geometry column. Similarly, literals of the `strdf:period` datatype are stored in a separate table with a GiST index on values of the `PERIOD` datatype, which is a primitive provided by PostgreSQL Temporal. The *Named Graph Translator* translates the temporal triples of stRDF to standard RDF triples following the named graphs approach as discussed in [2].

In the query engine, the optimizer and the evaluator have been implemented by modifying the corresponding components of Sesame to be able to evaluate spatial and temporal functions efficiently. The main idea is to push the evaluation of spatial and temporal extension function to the database level, so that the respective indices should be used to increase performance. The *stSPARQL to SPARQL 1.1 Translator* translates the temporal triple patterns to triple patterns using the named graph approach.

3 The Fire Monitoring Application

NOA operates an MSG/SEVIRI satellite acquisition station, and has developed a real-time fire hotspot detection service for effectively monitoring a fire-front. As soon as the images are acquired (every 5 or 15 minutes), they are stored as arrays in MonetDB. Then they are cropped and georeferenced using the SciQL query language and hotspots are detected producing shapefiles that are then translated into the stRDF data model. A number of refinement steps as stSPARQL update operations are then performed to increase the accuracy of the resulting products, i.e., false alarms and omission errors are detected and proper corrections are made, producing the final products of the processing chain. The stSPARQL

update operations that take place in the fire monitoring processing chain are described in [6]. In the fire monitoring application described in [6], only the refined results are eventually presented to the user. In the application that we present here, the user can also execute the refinement steps using the graphical user interface. Finally, we enrich the dataset that derives from the processing chain by combining it with the following two datasets (both compiled in the context of TELEIOS): (i) the dataset describing the coastline of Greece⁵, and (ii) the dataset describing the Greek environmental landscape⁶.

Strabon can also expose results in KML or GeoJSON and the results of each query posed can be depicted in a different layer of a map and overlay the retrieved data. This feature is very important for Earth Observation (EO) experts, as they can execute queries that retrieve EO data enriched with additional geospatial information from the Linked Open Data Cloud and visualize them in a map.

4 The Burn Scar Mapping application

The Burn Scar Mapping (BSM) is another application of the National Observatory of Athens. It involves the damage assessment using Landsat images, i.e. the estimation of the burned areas after wildfires. The products of the BSM service are in the form of shapefiles that contain information about the burned areas, i.e., their id, geometry, etc. We translate the shapefiles into stRDF graphs. The valid time dimension of the data model stRDF is also used to annotate triples to encode that an area has a respective land cover in a specific time period. The geospatial features of Strabon allow us to generate burned area maps automatically and enrich the displayed geometries with auxiliary information from other geospatial datasets or from the linked data cloud. The temporal features of Strabon allow us to express queries regarding the evolution of the land cover of an area that got burned at one or more time points of the time line. The final products of the BSM processing chain can be exploited into this direction, as they allow us to perform a time-series analysis since 1984.

5 Demonstration

The demonstration of the system is divided into two parts. The first part shows the fire monitoring application of NOA that uses the geospatial features of Strabon combined with the user-defined time dimension. Firstly, the service is initialized and the user can see a map depicting the current hotspots and then she can navigate into archive data and select hotspots from previous fire seasons and varying time ranges and display them in the map, as in Figure 2(a). Then she will execute the respective stSPARQL query and update operations that take part in the processing chain of the fire monitoring application.

The second part of the demonstration shows the burn scar mapping application of NOA that uses the geospatial features combined with the valid time

⁵ http://geo.linkedopendata.gr/coastline_gr/

⁶ <http://geo.linkedopendata.gr/corine/>



Fig. 2. (a) Fire monitoring application of NOA (b) Strabon endpoint

dimension of Strabon. Similar to the fire monitoring application, the user will be able to execute queries against Strabon asking for burned areas and then these areas will be displayed at a separate layer on the map. The dataset that will be used in this demo will contain the burned areas from the shapefiles provided by NOA combined with the CORINE land cover dataset. By this way, the user will be able to execute a query to retrieve the evolution of the land cover of an area that got burned at one or more points in the timeline. The user will be provided with a Strabon endpoint where she will be able to see and execute some pre-defined queries, as in Figure 2(b) and then write similar queries in the text area of the endpoint and execute them against the Strabon backend. For example, a user could pose a query to retrieve the previous land cover of burned area or the geometries of areas that were initially coniferous forests and then got burned.

References

1. Open Geospatial Consortium. OGC GeoSPARQL - A geographic query language for RDF data. OGC Candidate Implementation Standard (2012)
2. Bereta, K., Smeros, P., Koubarakis, M.: Representing and querying the valid time of triples for linked geospatial data. In: ESWC. LNCS, Springer (2013)
3. Cohn, A., Bennett, B., Gooday, J., Gotts, N.: Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *Geoinformatica* 1(3), 275–316 (1997)
4. Egenhofer, M.J.: A Formal Definition of Binary Topological Relationships. In: FODO. LNCS, vol. 367, pp. 457–472. Springer (1989)
5. Koubarakis, M., Kyzirakos, K.: Modeling and Querying Metadata in the Semantic Sensor Web: The Model stRDF and the Query Language stSPARQL. In: ESWC. LNCS, vol. 6088, pp. 425–439. Springer (2010)
6. Kyzirakos, K., Karpathiotakis, M., Garbis, G., Nikolaou, C., Bereta, K., Sioutis, M., Papoutsis, I., Herekakis, T., Mihail, D., Koubarakis, M., Kontoes, C.: Real Time Fire Monitoring Using Semantic Web and Linked Data Technologies. In: 11th International Semantic Web Conference. Boston, USA (2012)
7. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: A Semantic Geospatial DBMS. In: ISWC. LNCS, vol. 7649, pp. 295–311. Springer (2012)