Wildfire Monitoring Using Satellite Images, Ontologies and Linked Geospatial Data*

K. Kyzirakos[1], M. Karpathiotakis[1], G. Garbis[1], C. Nikolaou[1], K. Bereta[1], I. Papoutsis[2], T. Herekakis[2], D. Michail[3], M. Koubarakis[1], and C. Kontoes[2]

[1] National and Kapodistrian University of Athens
[2] National Observatory of Athens
[3] Harokopeio University of Athens

Abstract. Advances in remote sensing technologies have allowed us to send an ever-increasing number of satellites in orbit around Earth. As a result, Earth Observation data archives have been constantly increasing in size in the last few years, and have become a valuable source of data for many scientific and application domains. When EO data is coupled with other data sources many pionner applications can be developed. In this submission to the semantic web challenge we show how EO data, ontologies, and linked geospatial data can be combined for the development of a wildfire monitoring service that goes beyond applications currently deployed in various EO data centers. The service has been developed in the context of European project TELEIOS that faces head in the challenges of extracting knowledge from EO data, capturing this knowledge by semantic annotation encoded using EO ontologies, and combining these annotations with linked geospatial data to allow the development of interesting applications.

1 Introduction

TELEIOS is a recent European project that addresses the need for scalable access to petabytes of Earth Observation (EO) data and the discovery of knowledge that can be used in applications. To achieve this, TELEIOS builds on scientific databases, linked geospatial data, ontologies, and techniques for discovering knowledge from satellite images. Here we demonstrate the challenges and benefits of using ontologies and linked geospatial data in EO applications using a wildfire monitoring service as example.

The presented wildfire monitoring service significantly improves the wildfire monitoring service used until now by the National Observatory of Athens, a partner of TELEIOS. This service had been developed in the context of various GMES (Global Monitoring for Environmental and Secutiry) projects and is a good representative of state of the art work in this area so far. The service is available over the web[4] and has already been operational and used by decision

* This work has been cofunded by the FP7 project TELEIOS (257662) and the NSRF project SWeFS (180).

[4] `http://papos.space.noa.gr/fend_static/`

makers and emergency services in Greece during the summer of year 2012. This wildfire monitoring service is also presented in the Posters and Demonstration track of ISWC 2012 [4].

The front-end of the wildfire service, that we present, has been developed in the context of Sensor Web Fire Shield (SWeFS) research project funded by the Greek government by using software modules developed in TELEIOS.

The rest of the submission is organized as follows. Section 2 describes the workflow of the service and the utilization of the underlying components. Section 3 presents the datasets and the ontologies that are used in the fire monitoring service. Section 4 describes the operations that take place in order to increase the accuracy of the results. Section 5 presents some evaluation results and Section 6 concludes the submission. Appendix describes how the service addresses the minimal requirements and the additional desirable features of the Semantic Web Challenge 2012.

2 The Wildfire Monitoring Service

The wildfire monitoring service, that we have developed, relies on an MSG/SEVIRI acquisition station, operated by National Observatory of Athens (NOA). The station receives satellite images every 5-15 minutes. These images are stored in the state-of-the-art column-store DBMS MonetDB[5] and a fire detecting processing chain [6] is applied to them. The processing chain detects pixels where fire may exist (hotspots) and extracts shapefiles describing these hotspots. MonetDB provides SciQL [7], a new SQL-based query language for scientific applications with arrays as first-class citizens that has been developed in the context of TELEIOS. This allows us to store EO data (e.g., satellite images) in the database as arrays and express low level image processing (e.g., cropping, georeferencing) and image content analysis (e.g., pixel classification) in a high-level declarative language that provides efficient array manipulation primitives.

The products of this processing chain are encoded to stRDF. The model stRDF is an extension of the W3C standard RDF, that allows the representation of geospatial data that changes over time [2, 1]. stRDF is accompanied by stSPARQL, an extension of the query language SPARQL 1.1 for querying stRDF data. stRDF and stSPARQL use OGC standards (Well-Known Text and Geography Markup Language) for the representation of temporal and geospatial data and are implemented, in the context of TELEIOS, in Strabon[6], a fully implemented semantic geospatial DBMS that can be used to store linked geospatial data expressed in stRDF and query them using stSPARQL [5]. stSPARQL is essentially a subset of the recent OGC standard GeoSPARQL since it offers almost exact functionalities with the core, geometry extension and geometry topology extension of GeoSPARQL. Strabon supports this subset of GeoSPARQL as well.

In this wildfire monitoring service stRDF is used to represent satellite image metadata (e.g., time of acquisition, geographical coverage), knowledge extracted

[5] http://www.monetdb.org/

[6] http://www.strabon.di.uoa.gr/

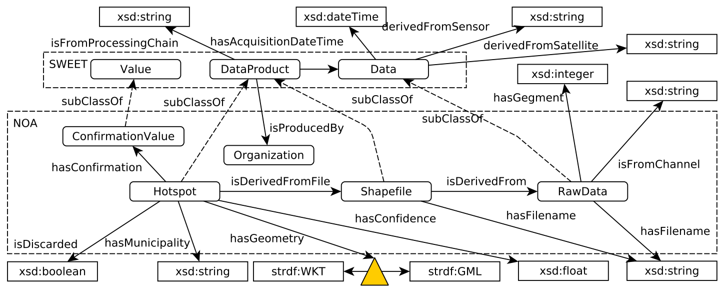

Fig. 1. The NOA ontology

from satellite images (e.g., a certain image region is a 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. By correlating detected hotspots with auxiliary data we associate more sophisticated information with them (e.g., municipalities that are affected by the fire) and we increase their accuracy (e.g., false alarms are detected and discarded).

Finally, the derived products are visualized in a map through a web application. The application enables real time wildfire monitoring while new satellite acquisitions arrive and are instantly processed. Additionally, a user can search the archive for a wildfire detected in the past and get a snapshot of it or watch (via simulation) how it proceeded. Information about hotspots (e.g., fire ignition time, nearby municipality) that are important for decision makers and crisis managers are also presented to the user.

3 Ontologies and Linked Geospatial Data

Apart from the hotspot data derived from the processing chain described in Section 2 the wildfire monitoring service also uses a variety of auxiliary geospatial data. Here we describe these datasets and how they are generated.

To be able to transform the detected hotspots into stRDF we developed the NOA ontology that describes the hotspot data. The NOA ontology (depicted in Figure 1) mainly consists of the classes RawData, Shapefile, and Hotspot which represent files with raw data (e.g., satellite images), ESRI shapefiles which are the outputs of the processing chain and hotspots which are extracted from the shapefiles, respectively. To achieve interoperability these classes have been defined as subclasses of corresponding classes of the SWEET ontology.

To enrich the aforementioned dataset of NOA we compiled in stRDF the following datasets.

Greek Landscape. The Corine Land Cover (CLC) project is an activity of the European Environment Agency (EEA) that collects data regarding the land cover of European countries and publish them as shapefiles. We developed an ontology that captures the hierarchical scheme that is adopted by this project

and we transformed the CLC shapefiles into stRDF according to this ontology.
Greek Coastline. This dataset originates from an ESRI shapefile, owned by NOA, describing the geometry of the coastline of Greece. For each area contained in this shapefile a unique URI is created and a spatial literal (which defines the geometry of the underlined area) is attributed to it.
Greek Administrative Geography. We developed an ontology that describes the administrative divisions of Greece (prefecture, municipality, district etc.). The ontology has been populated with relevant data (formated as CSV and SHP files) that are available in the Greek open government data portal[7].

3.1 Datasets from Linked Open Data Cloud

We enriched our dataset with well known datasets from Linked Open Data Cloud (LODC), such as OpenStreetMap (OSM) and GeoNames These datasets can be correlated with hotspot data to derive more information (e.g., find big roads that may stop a fire). In addition, these datasets cover a large variety of geospatial entities, ranging from fine-grained geometric objects (e.g., fire stations, hospitals) to coarser ones like countries. So one can exploit the ability of Strabon to expose data in KML or GeoJSON and create a map by posing queries on these datasets and overlaying the results.

4 Hotspots Semantic Refinement

The datasets described in Section 3 are mainly used to enhance the information captured by hotspot data and increase its accuracy. In this section we describe a series of refinement steps, using stSPARQL updates, that enrich hotspot data with information about nearby municipalities and increase its accuracy by detecting and correcting false positives or omission errors.

Notably, the queries described below are sophisticated update statements that exploit extensively the expressivity of SPARQL 1.1 and stSPARQL (e.g., GROUP BY, HAVING, aggregations, OPTIONAL) to cover the needs of the real time wildfire monitoring application of NOA.

Attribute Enrichment. Each hotspot is connected with a municipality where it is located, using the Greek Administrative Geography dataset. This is crucial information offered to decision makers and crisis managers for the optimal allocation of their firefighting resources. This refinement is performed by the following statement that inserts as a hotspot property the name of a municipality that spatially intersects with it.

```
INSERT { ?h gag:hasMunicipality ?municipality }
WHERE { SELECT ?h (SAMPLE(?mLabel) AS ?municipality)
  WHERE { ?h rdf:type noa:Hotspot; noa:hasGeometry ?hGeo;
          ?m rdf:type gag:Municipality; rdfs:label ?mLabel; strdf:hasGeometry ?mGeo.
          FILTER(strdf:intersects(?hGeo, ?mGeo)).}   GROUP BY ?h}
```

[7] http://geodata.gov.gr/geodata/

Consistency Refinement. The thematic consistency of the hotspots generated by the processing chain is achieved by the refinement step that correlates them with auxiliary geospatial data. This is done by a series of stSPARQL updates on the stRDF representation of the hotspots by taking into account relevant stRDF datasets from the ones presented above. The first step is to delete all hotspots that lie in the sea. Classification of pixels inside the sea as hotspots, is commonly encountered in wildfire scenarios near the coast. The hot smoke spreads above the sea, leading to misclassification of the corresponding pixels. This operation is performed by the following stSPARQL update that marks as discarded every retrieved hotspot. We are using the query applied previously: if a hotspot does not spatially intersect a municipality, it lies in the sea, so it is discarded.

```
INSERT {?h noa:isDiscarded "1"^^xsd:int}
WHERE { ?h rdf:type noa:Hotspot.
        OPTIONAL {?h gag:hasMunicipality ?municipality}.
        FILTER(!bound(?municipality)).}
```

In a similar way, hotspots that are in the mainland, but lie in non-consistent land cover areas are also discarded. Using the Greek Landscape dataset, the non-consistent classes are defined as i) artificial surfaces, ii) agricultural areas (arable land & permanent crops), and iii) wetlands and water bodies. This operation is performed by the following stSPARQL update.

```
INSERT {?h noa:isDiscarded "1"^^xsd:int}
WHERE { SELECT ?h
        WHERE { ?h rdf:type noa:Hotspot; noa:hasGeometry ?hGeo.
                ?a rdf:type clc:ExcludeArea; clc:hasGeometry ?aGeo.
                FILTER(strdf:mbbIntersects(?hGeo,?aGeo)).}
        GROUP BY ?h ?hGeo HAVING strdf:contains(strdf:union(?aGeo),?hGeo)}
```

To ensure the product visualization consistency, we also utilize the Greek Coastline dataset and keep only the part of a hotspot polygon that lies in land, and eliminate the part that lies in the sea.

Temporal Persistence. The fire detecting processing chain (Section 2) identifies hotspot pixels and marks them either as "potential fires" with a confidence level of 0.5 or as "certain fires" with a confidence level 1.0. The algorithm is based on a series of spectral tests with some thresholds. Setting appropriately these thresholds is the outcome of a trade-off between omission errors and false alarms. In certain scenarios, this leads to a phenomenon described as "Christmas tree effect", where some hotspots appear for the first time, in the next timestamp they disappear, then they re-appear again, and so on. To avoid this effect we also examine the temporal persistence of each hotspot and decide on whether to insert a new hotspot on the database or not.

5 Evaluation of Performance

In order to provide a real time wildfire monitoring application it is important that the refinement operations described in Section 4 are performed at reasonable time. We have carried out several experiments for the refinement operations

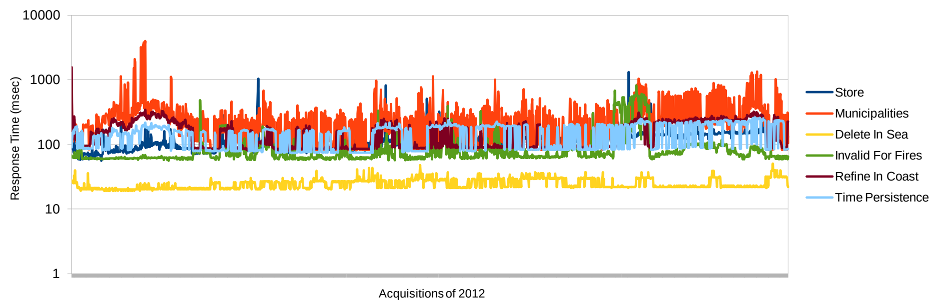

Fig. 2. Response time of store and refinement operations for each acquisition of 2012

on datasets containing hotspots detected during the fire seasons of the years 2007, 2008, 2010, 2011, and 2012. Figure 2 shows the results of the experiment using the acquisitions of year 2012. We observe that all operations are executed efficiently, mostly in less than a second, except for the operation of associating detected hotspots with the municipality they belong to. This operation is labeled as "Municipalities" in Figure 2. Although in most cases the query processing time is less than two seconds, there are cases where the operation needs four seconds to be completed, but even then the performance of the system is satisfactory.

As described previously, the refinement steps are expressed as complicated stSPARQL update statements (i.e., using aggregates, functions, etc.). Their efficient execution is mainly owed to Strabon's high performance and scalability, even in much larger datasets than the ones used in this application (experiments with up to 500 millions triples have been performed). Strabon is one of the most rich, in terms of functionality, and efficient RDF stores available today and in many cases outperforms related competitor systems, as documented in [5].

6 Conclusion

In this contribution to the Semantic Web Challenge 2012, we present the wildfire monitoring service that we have developed in TELEIOS and we highlight the contributions of the ontologies and linked geospatial data. New ontologies and datasets encoded to stRDF have been created for this purpose, and have been combined with existing open datasets. Refinement operations, encoded to stSPARQL, are integrated in the workflow of NOA, producing validated results in real time. The processing chain of the service is executed efficiently on top of Strabon and MonetDB, and the user is able to access real time and archived hotspot data via a user-friendly interface. The same principles can be adopted to other aspects of wildfire management or other applications. For example, in the context of SWeFS we are currently extending the service with fire prediction models. This requires creating and integrating more linked geospatial data about weather, human activities and flammability of specific regions.

References

1. Koubarakis, M., Karpathiotakis, M., Kyzirakos, K., Nikolaou, C., Sioutis, M.: Data Models and Query Languages for Linked Geospatial Data. In: Reasoning Web. Semantic Technologies for Advanced Query Answering. LNCS, Springer (2012)
2. Koubarakis, M., Kyzirakos, K.: Modeling and querying metadata in the semantic sensor web: The model stRDF and the query language stSPARQL. In: ESWC (2010)
3. Koubarakis, M., Sioutis, M., Kyzirakos, K., Karpathiotakis, M., Nikolaou, C., Vassos, S., Garbis, G., Bereta, K., Dumitru, O.C., Molina, D.E., Molch, K., Schwarz, G., Datcu, M.: Building Virtual Earth Observatories using Ontologies, Linked Geospatial Data and Knowledge Discovery Algorithms. In: ODBASE (Sep 2012)
4. Kyzirakos, K., Karpathiotakis, M., Garbis, G., Nikolaou, C., Bereta, K., Sioutis, M., Papoutsis, I., Herekakis, T., Michail, D., Koubarakis, M., Kontoes, C.: Real Time Fire Monitoring Using Semantic Web and Linked Data Technologies. In: ISWC (2012)
5. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: A Semantic Geospatial DBMS. In: Proceedings of the 11th International Semantic Web Conference (2012)
6. Sifakis, N., Iossifidis, C., Kontoes, C., Keramitsoglou, I.: Wildfire Detection and Tracking over Greece Using MSG-SEVIRI Satellite Data. Remote Sensing (2011)
7. Zhang, Y., Kersten, M.L., Ivanova, M., Nes, N.: SciQL: bridging the gap between science and relational DBMS. In: IDEAS (2011)

A Appendix: Compliance with Semantic Web Challenge Requirements

A.1 Minimal requirements

End-user application. The service has a web based interface tailored for decision makers and crisis managers interested in real time wildfire monitoring.

Information sources. The service uses a variety of information sources under different ownership and control. The core of our dataset contains fire related products extracted from satellite images owned by NOA enriched with data published by the European Environment Agency and by the Greek government and with data published in LODC. Because of various ownership the data is highly heterogeneous. Some of the datasets are complex with high hierarchical structure like Greek Administrative Geography while other (e.g., Greek Coastline) are simpler and have a flat structure. Used data comprises a representative amount of spatial information about Greece. It contains approximately 1 million triples that describe 10,000 different elements.

Data meaning. Data is represented using stRDF to capture its spatial extent. Also, OWL ontologies are used to describe the structure of the data. Today EO practices rely on processing EO data as it is. We propose using Semantic Web technologies for automatically comparing EO data with auxiliary data and deriving more complex kinds of information (e.g., fire in sea) that basic EO practices (e.g., hotspots detection) can not capture.

A.2 Additional Desirable Features

Web interface. While new hotspots are detected the application animates the evolution of a fire-front along with useful auxiliary information (e.g., affected municipalities). Refinement operations are hidden from the end-user for simplicity. Also, a user-friendly option is offered to retrieve and watch past wildfires.

Scalability. Our approach scales to other EO applications and this is what we do with other use cases of TELEIOS e.g., in [3]. In addition, the scalability of our application is guaranteed since its performance is entirely dependent on the systems, Strabon and MonetDB, which have been shown to be scalable.

Evaluation. Rigorous evaluations have proved that the stSPARQL operations performing hotspot refinement are executed efficiently. The application was operational since the beginning of August 2012 and detected and presented successfully the wildfires that burst during that month in real time. By this time, the application had been available to the user community for receiving feedback.

Novelty. We applied Semantic Web technologies to encode EO products produced from satellite images, integrate such products with auxiliary spatial information in order to increase the accuracy and enrich the generated information. To the best of our knowledge, this is the only operational fire monitoring service that uses semantic web technologies and linked open geospatial data, and one of very few EO applications that do the same.

Functionality. The application permits the real time monitoring of a fire-front. Additionally, the depicted hotspots are refined to be consistent with the areas where they lie in.

Commercial potential and/or large existing user base. The application assists the management of an important environmental issue. Many countries and organizations are interested in good solutions for environment management (and especially wildfire monitoring) because it can drastically affect the environment and human life. A lot of organizations internationally are interested in such services.

Contextual information. Hotspots confidence (initially generated from hotspot detection) is refined using stSPARQL queries.

Multimedia. Satellite images play a crucial role in the service as long as they constitute the source data for deriving hotspots data.

Dynamic data. A satellite station is used to acquire images every 5-15 minutes to provide real time fire-front monitoring. Static information are used to refine and enrich information about detected hotspots.

Accuracy. The accuracy of detected hotspots are improved by correlating them with auxiliary information about the background. False fires and omission errors are detected and corrected.

Multilinguality and accessibility on a range of devices. The application GUI is implemented only in English but presented data can be in any language (currently Greek) if data is UTF-8 formated. The application is based in Javascript and can be accessed by any well known web browser.