Semantics for spatio-temporal "smart queries"

Benjamin Harbelot, Helbert Arenas and Christophe Cruz
Laboratoire Le2i, UMR-6302 CNRS,Departement Informatique , University of Burgundy, 7 Boulevard Docteur Petitjean, 21078 Dijon, France
{*benjamin.harbelot,helbert.arenas*} *@checksem.fr, christophe.cruz@u-bourgogne.fr*

Keywords: Spatial Data Infrastructure, Semantics, Spatio-temporal

Abstract: There is a demand to enhance the meaning of objects and their relationships in the geospatial domain.This need has led to a new field of research called geosemantics, which takes advantage of state of art databases and inference mechanism of ontologies. Current geosemantic proposals involve mixed solutions using spatial databases to store geographic objects and a separate ontology to provide semantic capabilities to the system. However making a simple query involves requesting information from both components which is problematic. In our previous work we proposed a functional standalone based architecture, that uses SWRL to solve spatio-temporal problems.In this paper we present an enhancement of our previous work using a service oriented architecture. The new architecture makes us of triplestores and data from WFSs and SOSs making it more flexible.

1 INTRODUCTION

Nowadays available services such as OGC's WMS, WFS or WCS provide a great amount of spatial data. This information allow scientist to perform complex analysis. However most of these data sources operate only at a syntactic level, lacking the necessary semantics for richer spatio-temporal applications. In (Harbelot et al., 2012) we introduced the Continuum model to manage the semantics of the evolution of elements over time and space. This research led to the implementation of temporal and spatial built-ins that allow us inference capabilities within an ontology. The SWRL built-ins perform temporal operations and make use of a PostGIS database to have access to spatial operators. The results of these operations are later added to the ontology enriching the knowledge base. The main limitation of this system is that it runs on a local computer. In this article we propose an extension to our previous work making our implementation a component of an information sharing community. By allowing our implementation to access a data sharing community we are able to perform complex analysis by combining multiple and heterogeneous data sources. (Goodwin, 2005) labelled this type of request as *"smart queries"* . To be able to achieve this goal it is necessary to define formal semantics to the different data sources and processes, which can be done using ontologies. Using ontologies in the field of Geographic Information Systems allow us to combine heterogeneous data sources, and obtain new non explicit knowledge using inference mechanisms. In this paper we present an architecture that allow us to perform *"spatio-temporal smart queries"* using OGC standards. In section 2 we describe previous research on this field. On section 3 we describe the architecture we are currently implementing, later on section 4 we describe how the system would work using an example case scenario and finally in section 5 we present our conclusions.

2 RELATED RESEARCH

Spatial Data Infrastructure (SDI) is a term first introduced by the U.S. National Research Council in 1993. It refers to a set of technologies, policies and agreements designed to allow the sharing of spatial information and resources between institutions (ESRI, 2010). The Spatial Data Infrastructure has a service oriented architecture. In such infrastructure, functionalities such as storage and data search are carried out through Web services. The typical work flow involves 1) The discovery of a data source, 2) The download of relevant geo spatial data, 3) Applying the appropriate analytical methods and 4) The visualization of the results on a suitable map.

The basics SDI software components are: 1) Client sofwares, 2) Catalogue services, 3) Spatial data services, 4) Processing services 5) Spatial data repositories, 6) GIS to create/update the spatial data.

Additionally a set of technical standards is necessary to allow the interoperability between the different software components. Currently in the field of geo information science the standards are defined by the Open Geospatial Consortium (OGC). This entity is an international industry consortium, composed by 481 companies, government agencies and academic institutions (OGC, 2012). Among the OGC standards we have: WMS (for map service providers), WFS (for vector feature providers) and GML (for data formatting). Another very important OGC standard is CSW which allows the interaction with one or more catalogues of spatial resources. Spatial resources can be map data of the OGC Web services (WMS, WFS, WCS), spatial data repositories or sensors (SOS). CSWs have two main capabilities:1) Allow the discovery of data and 2) Metadata management (Nebert et al., 2007). A CSW is able to search for spatial data or web services using an open criteria (i.e. free text as a search in a search engine) or on more specific criteria (title, coordinate system, data type, etc.). Another capability of a CSW is the synchronization of content with another catalogue. A CSW can also manage metadata, by performing addition, modification or deletion operations. A very interesting capability of the CSW is the harvesting operation. Using this operation a CSW can address a data provider and retrieve the description of the available information by accessing the *GetCapabilities* method. Harvesting can also be used to collect metadata from a remote catalog and then store it locally to allow faster access. The harvesting operation is performed periodically and allow the local metadata and the remote catalogue metadata to be synchronized.

All OGC services implement the *OWS Common* which describe basic parameters and data structures used in the request or response from web service operations. The standardization proposed through OWS Common serves as the ground base for the interoperability of OGC web services (Whiteside, 2005). The implementation of interoperability within a SDI present several semantic challenges as explained in (Janowicz et al., 2010). Most semantic problems arise due to the lack of suitable content descriptions. In order to design a SDI with interoperability capabilities is necessary to add proper semantics to data, metadata and services.

Currently most of the available information on the web is provided as syntactic blocks, which require human experts to determine their semantics. The needed semantics can be provided by ontologies which would allow software inference mechanisms to reason with it. In the Artificial Intelligence field, an ontology is an engineering mechanism, constituted by a vocabulary and a set of assumptions regarding this vocabulary. This information is specified using a first order logic theory. Using an ontology it is possible to specify a concept, to define a number of restrictions that apply to a certain concept and create a hierarchy of concepts with subsumption relationships among them (Guarino, 1998).

A common approach to represent the knowledge stored in an ontology is using Description Logics (DL). An ontology has two main parts, the T-Box and the A-Box. The T-Box contains the definition of *concepts* which represent sets of individual entities, and *roles*, which are binary relationships between individuals. The A-Box contains a set of assertions about individuals, using the *concepts* and *roles* defined in the T-Box. Typical reasoning with a T-Box involves 1) The analysis of *concept* descriptions, determining if a definition is not contradictory, and 2) Comparison between two concepts to determine if one of them is more general than the other (subsumtion analysis). Typical analysis using a A-Box determines if a set of assertions is valid, or if a certain individual is an instance of a given *concept* (Baader and Nutt, 2003).

Our goal is to design the necessary mechanisms to perform *"smart queries"*. This term was introduced by (Goodwin, 2005), it refers to analysis that integrates heterogeneous data sources to achieve complex tasks. In Goodwin (2005), the author presents an example in which two data sources are used, the first contains information about the location of pollutants while the second provides information about water bodies. An analysis of the later provides information about the topological relations between water bodies (For instance *connectedTo*). Then when combining these two data sources we could identify water bodies that overlap with known pollutants. Then by using the topological relations between waterbodies we could infer which elements are potentially at risk of becoming contaminated due to being connected to already polluted ones (Goodwin, 2005). Using Description Logics we can define the concept *Area*, then define the relations $connectedTo$ and $contains$. Then the formal definition of a *RiskArea* is:

$RiskArea \subseteq Area \sqcap \exists connectedTo.(Area \sqcap \exists contains.Pollutant)$

$RiskArea$ is an $Area$ which has a relation $connectedTo$ with another $Area$, which has the value $Pollutant$ for the relation $contains$.

The ideas introduced by (Goodwin, 2005) are later further developed by (Kammersell and Dean, 2007). In this research the authors identify 3 main limitations for traditional queries:1) In keyword and database searches, the users must be aware of the data source vocabulary and syntax. 2) There are problems that involve requesting information from diverse data sources, forcing the users to perform individual queries to each one of them. 3) Traditional queries can not express complex ideas like preferences. In order to overcome these limitations Kammersell and Dean (2007) propose the use of a *Conceptual Search*. A *Conceptual Search* combines smaller semantic queries allowing the formation of complex data requests from a variety of data sources (Kammersell and Dean, 2007). Kammersell and Dean propose the creation of a layer that translates the user's query formulated in OWL to a WFS XML request, and later do the inverse process with the results produced by the WFS to add the information into the user's ontology.

The implementation of a knowledge system that deals with large quantities of geospatial knowledge demands a suitable data storage. Currently most of the spatial data in traditional systems are stored in relational databases. RDBMS can handle large datasets with the help of efficient spatial indexing and operators. However relational databases lack inference mechanisms that can be applied to ontologies. Triplestores are designed to handle large ontologies, however availability of spatial capabilities is not the norm. In (Battle and Kolas, 2012) the authors review a number of triple stores implementations, indicating their strong and weak points and later introduce Parliament a triple store with spatial indexing capabilities that supports GeoSPARQL which is an emerging standard from OGC (Battle and Kolas, 2012). The goal of GeoSPARQL is to provide a consistent logical way to express geo spatial Semantic Web data using RDF. GeoSPARQL defines a small ontology for representing features, geometries and a number of query predicates and functions. This ontology closely follows existing standards from the OGC. GeoSPARQL has three components: 1) GeoSPARQL ontology, 2) GeoSPARQL relationships, and 3) Query Transformation rules.

3 PROPOSED IMPLEMENTATION

Currently metadata of datasets and services are stored in CSW services. We see the information stored in CSWs as the syntactic building blocks necessary to create a semantic capable SDI. From the CSW it is possible to know which of WFSs or SOSs contains a given information and how can it be accessed (data layer name, format, spatial representation). We propose the mapping of this information into the T-Box of an ontology. This process would be done with the help of a human expert. A *concept* on the ontology could refer to features stored in several data layers in more than one data server. It is also possible that the information from one layer in a WFS could contain features corresponding to more than one *concept* in the ontology.

Figure 1 depicts how this process would occur between the CSW data and an ontology. The picture represents on the right side the information in a CSW corresponding to four WFS each containing more than one data layers. For example, the *WFS Census* contains the layers *UrbanCenters*, *Cities*,*Provinces*,*Departments*. Each layer contains the geographic representation of a set of features and their associated attributes. With the help of the human expert the relevant data layers will be mapped into one or more *concepts* in the ontology.

The left side of figure 1 represents the *concepts* of a hypothetical ontology, presented only for the purposes of the example. In the ontology, the base object is called $GeoObject$. It has several specifications. For the specification of *concepts* and their relationships we will use the notation proposed by (Baader and Nutt, 2003)). Using this notation the specification for the *concept Business* is:

$Business \equiv \overline{Business} \sqcap GeoObject$

where $\overline{Business}$ refers to all the characteristics that distinguish a *Business* from other elements of the class $GeoObject$.

Instances for the class *Busines* can be obtained from the data layers `GasStation` from the `WFS Transport`, and `Banks` and `Businesses` from `WFS Municipal`. The information for the instances is obtained from WFS GetFeature requests. The resulting GML is translated into OWL using XLST.

In the example ontology, the classes *Restaurant* and *Factory* are specifications of the class *Business*.

$Restaurant \equiv Business \sqcap \forall HasType.RESTAURANT$

$Factory \equiv Business \sqcap \forall HasType.FACTORY$

Instances for these classes can not be obtained straightforward from any data layer without a previous filtering. There are two alternatives for this process. 1) It can be done within the ontology, ie. a Restaurant is an instance of the class *Business* that holds a suitable criteria ($\forall HasType.RESTAURANT$). or 2) The information for the instances can be obtained from the WFS using a filter option. The use of filters allows a great amount of freedom for the definition of classes in the ontology. For instance we can define the concept $Restaurant\_on\_District\_1$. This concept refers to restaurants located within the

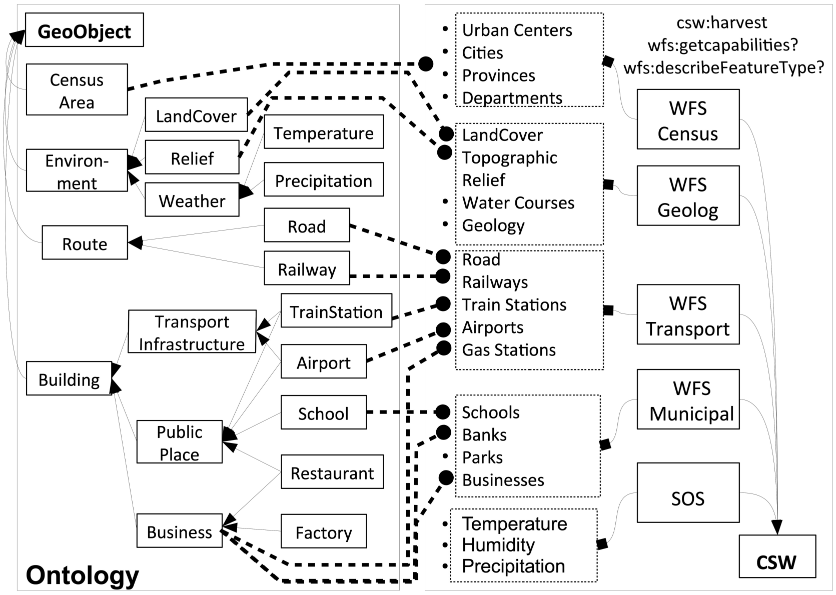

Figure 1: CSW data mapping to an ontology

boundaries of an element called District 1:

$Restaurant\_on\_District\_1 \equiv Restaurant \sqcap Within(Restaurant, District\_1)$

It is possible to populate this concept by applying spatial and alphanumeric filters to the request from `WFS Municipal`:

On (Harbelot et al., 2012) we have introduced spatial SWRL built-ins that could be adapted to perform this operation within an ontology. The SWRL built-in could be used to test if an instance of the class *Restaurant* is within the boundary of District 1. An alternative to this process is to directly populate the class *Restaurant_on_District_1* using WFS alphanumeric and spatial filters. We could obtain the spatial representation of District 1 and then request from `WFS Municipal`, features from layer `Business` that are Restaurants and are located within the boundaries of District 1.

```
http://WFS_Municipal?
service=WFS&
VERSION=1.0.0&
typeName=Business&
request=GetFeature&
filter=
<filter>
<PropertyIsEqualTo>
<PropertyName>Business_Type</PropertyName>
<Literal>Restaurant</Literal>
</PropertyIsEqualTo>
<Within>
<PropertyName>the_geom</PropertyName>
<gml:Polygon>
<gml:outerBoundaryIs>
<gml:LinearRing>
<gml:coordinates>
40.78482,-73.98146 40.78721,-73.98146
40.78721,-73.97716 40.78482,-73.98146
</gml:coordinates>
</gml:LinearRing>
</gml:outerBoundaryIs>
</gml:Polygon>
</Within>
</filter>
```

The first part of the filter selects only features for which the value of the attribute `Business_Type` is `Restaurant` while the second selects features located within the defined boundary.

In the example ontology, the class *Precipitation*, is a specification of the class *Weather*. None of the available data layers can be mapped directly into the class *Precipitation*. However a human expert might suggest that the information stored from a SOS that stores weather measurements can be further processed to obtain isolines or to classify the values of the observations in order to create a set of polygons.

In our proposed architecture the ontology resides on a triple store located in a web server accessible to users though the web using HTTP protocol. Once the T-Box *concepts* are defined, the A-Box is populated by translating the T-Box vocabulary into `wfs:GetFeature` or `sos:getObservation` requests with the appropriate filters. The response from the services is later translated from GML into OWL/RDF and saved in the triple store. Further processing and analysis is done in the ontology using customized built-ins and GeoSPARQL. Figure 2 depicts the system architecture. In previous work we have already

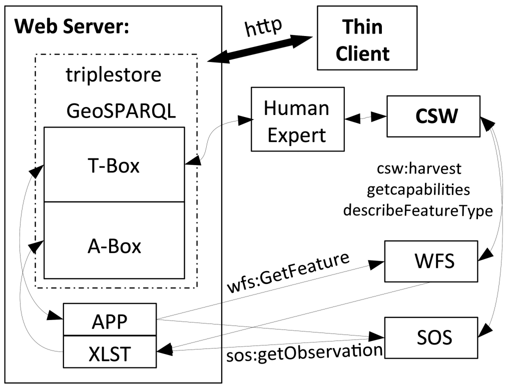

Figure 2: Proposed architecture

implemented built-ins for spatial and temporal inference, that would be re-used with little in the new implementation.

4 USE CASE SCENARIO

For the use case example scenario, lets suppose that due to extreme heavy rain, a significant part of a number of provinces is currently flooded. The task is to identify within the affected provinces, potential helicopter landing areas, these areas need to be unaffected by the flood and with good weather conditions for the next 24 hours.

First we construct a new concept that defines to the study area:

$FloodedProvince \equiv CensusArea \sqcap HasReported.FLOOD$

Second, we identify the unaffected zones within the study area:

$DryArea = \{a \in GeoObject | p \in FloodedProvince \wedge q \in FloodedArea \rightarrow a = geof : difference(q, p)\}$

Third, we identify areas with good weather conditions for the next 24 hours, which is the result of applying a alphanumeric filter plus a temporal operator. In this example we use the temporal operator *During*. Operators of this nature have been introduced in (Harbelot et al., 2012), in which we implemented SWRL operators based on Allen interval relationships.

$ClearSkyArea \equiv Precipitation \sqcap (\forall HasPrecipitation.LITTLE\_OR\_NO\_RAIN \sqcap (During(\forall HasTime.TIME, NEXT\_24\_HOURS))$

Finally, we identify the areas with good weather conditions and unaffected by the flood within our study area.

$PotentialLandingArea = \{b \in GeoObject | a \in DryArea \wedge r \in ClearSkyArea \rightarrow b = geof : intersection(a, r)\}$

We propose the use of already implemented GeoSPARQL operators. Among others GeoSPARQL supports: $geof : distance$,$geof : buffer$, $geof : convexHull$, $geof : intersection$,$geof : union$,$geof : difference$,$geof : symDifference$,$geof : envelope$,$geof : boundary$, $geof : getsrid$.

5 CONCLUSIONS

In this paper we introduce an architecture capable to store and handle data and and query multiple services like WFS's and SOS's. The proposed architecture is integrated in an Spatial Data Infrastructure taking advantage of existing OGC standards as a way to obtain data and perform complex spatial and temporal queries using diverse and heterogeneous data sources. The semantic SDI is able to provide a semantic, spatial and temporal analysis presented in the use case scenario.

Currently, the system needs a human expert to map information between ontology and catalog of metadata. Further research is required to provide an automatic (or semi-automatic) mapping. Moreover it should be interesting to connect mobile devices to the system opening the field of decision support system.

ACKNOWLEDGEMENTS

This research is supported by the Direction Général de l'Armement, see:

http://www.defense.gouv.fr/dga/

REFERENCES

Baader, F. and Nutt, W. (2003). The description logic handbook. chapter Basic description logics, pages 43–95. Cambridge University Press, New York, NY, USA.

Battle, R. and Kolas, D. (2012). Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*.

ESRI (2010). GIS Best Practices: Spatial Data Infrastructure (SDI).

Goodwin, J. (2005). What have ontologies ever done for us - potential applications at a national mapping agency. In *in OWL: Experiences and Directions (OWLED)*.

Guarino, N. (1998). Formal ontology and information systems. In *Proceedings of FOIS 98*, pages 3–15.

Harbelot, B., Arenas, H., and Cruz, C. (2012). A model and a query language to analyse semantico-spatio-temporal data. In *Subtmitted to ICDE 2013*.

Janowicz, K., Schade, S., Broring, A., Kebler, C., Maue, P., and Stasch, C. (2010). Semantic enablement for spatial data infrastructures. *Transactions in GIS*, 14(2):111–129.

Kammersell, W. and Dean, M. (2007). Conceptual search: Incorporating geospatial data into semantic queries. In Scharl, A. and Tochtermann, K., editors, *The Geospatial Web*, Advanced Information and Knowledge Processing, pages 47–54. Springer London. 10.1007/978-1-84628-827-2_5.

Nebert, D., Whiteside, A., and Vetranos, P. (2007). Catalogue services specification.

OGC (2012). OGC Institutional Web Site.

Whiteside, A. (2005). Web services common specification.