Integration of Spatial XML Documents with RDF

J. E. Córcoles, P. González, V. López Jaquero

LoUISE: Laboratory on User Interface & Software Engineering.
Universidad de Castilla-La Mancha
02071, Albacete (Spain)
{corcoles, pgonzalez}@info-ab.uclm.es

Abstract: The fact that GML is an XML encoding allows it to be queried. In order to query a GML document we have designed a query language over GML/XML enriched with spatial operators. This query language has an underlying data model and algebra that supplies the semantics of the query language. In this paper, we propose an approach for integrating Geospatial data on the Web, stored in GML documents, using our Spatial Query language. This approach has been inspired by Web Portals technology for retrieving Geospatial information. We have used ontologies to solve the semantic heterogeneity between different GML documents.

1 Introduction

Actually, a rich domain that requires special attention is the Semantics of Geospatial Information [1]. The enormous variety of encoding of geospatial semantics makes it particularly challenging to process requests for geospatial information. Work led by the OpenGIS Consortium [2] addressed some basic issues, primarily related to the geometry of geospatial features. The Geography Markup Language (GML) provides a syntactic approach to encoding geospatial information through a language in which symbols need to be interpreted by users, because associated behaviour is not accounted for.

The fact that GML is an XML encoding allows it to be queried. In order to query a GML document we have developed a query language over GML/XML enriched with spatial operators [3]. This query language has an underlying data model and algebra that supplies the semantics of the query language [4].

This preliminary work is the base for developing a Web environment that allows spatial queries over GML on the Web. At this moment, querying spatial XML documents (GML) requires knowledge of the DTD (XMLSchema). In addition, if two GML documents have semantically the same information but are represented by different DTDs, it is necessary to manually make two queries, one for each document. Thus, the user has to make two queries to select the same semantic information, and must know the DTD offered by each GML document. In order to solve this problem, it is necessary to implement some approach that allows querying of the common knowledge and interests that people share within their community. In this way, the

user can apply the same query and the same structure to query several GML documents that satisfy, either fully or partially, this general structure. This common knowledge is Ontologies and it is used in this paper for solving the semantic heterogeneity between Spatial XML documents.

In this paper, we propose an approach for integrating Geospatial data (expressed with GML documents) on the Web. We focus this approach in the definition of a mapping between the ontology and the DTD/Schemas. Our Spatial Query language over spatial XML documents is used in this approach like a canonical language that allows: (i) a single query language that the users can express their queries in, and (ii) a single query language for querying each GML document in the same way.

This paper is organized as follows: Section 2 briefly describes the catalog definition to facilitate the mapping between the general query (over de ontology) and concrete query over each GML document. The conclusions and future work are presented in Section 3.

2 Catalog Structure

The Catalog has been defined using the Resource Description Framework (RDF) standard [5][6] proposed by W3C in order to facilitate the creation and exchange of resource descriptions between Community Webs.
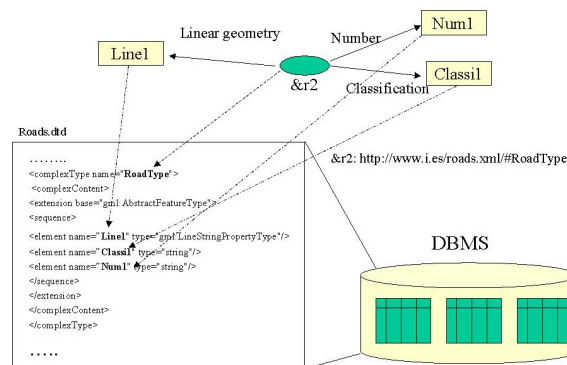

Fig. 1. Relation between ontology and Local Schemas

In the RDF model, the resources are parts of a DTD that contains the schema offered for each site. A simple resource may be a simple element or attribute in a DTD. A complex resource is a complex element in a DTD. The simple resources are represented by Literals in a RDF schema and the complex resources are represented by RDF. The RDF establishes a correspondence between the ontology and parts of the schema in each site (local schemas). With this information, it is possible to translate a

query expressed in an ontology into a query expressed in terms of the different schemas.

Figure 1 shows an example. We have used the graphical notation used in [6]. The resource &r2 defines an element *RoadType* offered by the site *http://www.infob.ulm.es/roads.xml*. The properties *LineGeometry*, *number* and *classification* are related to the elements *Line1*, *num1* and *Classi1* in the Resource *TypeRoad*. The DTD offered defines, partially or fully, the Schema of the DBMS.

In Figure 2, an example of the Portal Schema and its instances is shown. The example has been obtained from the specification documents of GML by OpenGIS [GML01]. Due to RDF's capability for adding new feature and geometry types in a clear and formal manner [7], this example has been carried out extending the geospatial ontology defined by OpenGIS, where the class (Geometry, LineString, etc) and properties (*coordinates*, *Polygonmember*, etc) are defined.
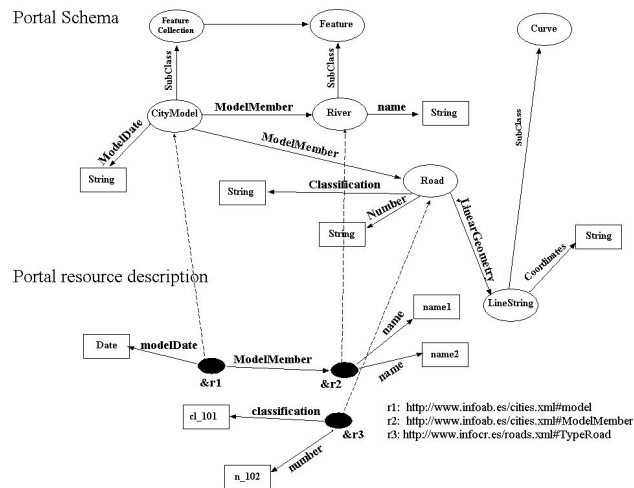

Fig. 2. Example of Catalog

The *cityModel* contains a property *modelDate* and *modelMember* (*River* or *Road*). A *River* is defined by a *name*, and the *Roads* are defined by *Classification*, *Number* and *LinearGeometry* (string of coordinates). *River* and *Road* are subclasses of the class Feature. *CityModel* and *LineString* are subclasses of the class *FeatureCollection* and *Curve* respectively. These classes are defined in the Geospatial ontology by [7]. In this RDF schema the properties inherited have not been included for the sake of simplicity. Except for the complex elements (*Road*, *LineString*, etc) the properties have a literal Range. Furthermore, the scope of the declarations is determined by the corresponding *namespace* of each schema, e.g. Ns1 (http://www.opengis.org/gml#). For the sake of simplicity, we will henceforth omit the namespaces prefixing in class and property names. In the Portal resources description the instances of the model are shown.

The translation of the general query expressed in terms of the Portal Schema, into queries expressed in terms of the sites is carried out querying the Portal Catalog using

RQL [8]. RQL is a declarative query language for RDF. It is defined by a set of basic queries and iterators which can be used to build new ones through functional composition.

3 Conclusions and Future Work

In this paper, we have proposed an approach for integrating Geospatial data (expressed with GML documents) on the Web. We focus this approach in the definition of a mapping between the ontology and the DTD/Schemas using RDF(S). Our Spatial Query language over spatial XML documents is used in this approach like a canonical language

With this approach it is possible to use a single spatial query language for GML documents without knowing the exact structure of each document. The main disadvantage of this approach is that each site has to upload the RDF model to be stored in the Catalog. The automatic generation of the catalog is a matter we intend to deal with in our future work.

References

1. Egenhofer J. Toward the Semantic Geospatial Web. ACM-GIS 2002. 10th ACM International Symposium on Advances in Geographic Information Systems. McLean (USA). 2002.
2. Open GIS Consortium, Inc. OpenGIS: Simple Features Specification For SQL Revision 1.1 OpenGIS 99-049 Release. 1999.
3. Córcoles, J. and González P. A Specification of a Spatial Query Language over GML. ACM-GIS 2001. 9th ACM International Symposium on Advances in Geographic Information Systems. Atlanta (USA). 2001
4. Córcoles, J. and González, P. A spatial query language over XML documents. Fifth IASTED International Conference on Software Engineering and Applications (SEA). Los Angeles, USA. pp.1-6. 2001
5. Brickley, D. and Guha, R.V. Resource Description Framework (RDF) Schema Speci_cation 1.0, W3C Candidate Recommendation. Technical Report CR-rdf-schema-20000327, W3C, March 2000. Available at http://www.w3.org/TR/rdf-schema.
6. Lassila O. and Swick, R. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, February 1999. Available at http://www.w3.org/TR/REC-rdf-syntax. 1999
7. OpenGIS. Geography Markup Language (GML) v2.0. Document Number: 01-029. Http://www.opengis.net/gml/01-029/GML2.html. 2001
8. Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D., Scholl, M. RQL: A Declarative Query Language for RDF,WWW'02.Hawaii, USA.2002.