VirGIS data in Semantic Web environment

Zdeňka Linková, Radim Nedbal

linkova@cs.cas.cz, nedbal@cs.cas.cz
Ústav informatiky, Akademie věd ČR,
Pod Vodárenskou věží 2, 182 07 Praha 8

Abstract. A crucial point in automated data processing is the way in which the data are expressed. One possibility is to employ existing features of the Semantic Web - ontologies. Ontologies play an important role in a knowledge representation.
The aim of the research presented in this paper is to provide more automated VirGIS system. VirGIS is an integration system that works with GIS (Geographical Information Systems) data. As a first step of our research, we describe its data using common Semantic Web techniques and build a VirGIS ontology.

1 Introduction

Today's world is a world of information. Everything depends on information, whether science progress or business success. Expansion of World Wide Web has brought better accessibility to information sources. However, in the same time, the big amount of different formats, data heterogeneity, and machine unreadability of this data have caused many problems. Data features make automated processing difficult. Exactly from this base rises the idea of the Semantic Web [1]. It considers data to go along with their meanings. An addition of semantics would make data machine readable and understandable. The automation could be easier.

A crucial point in automated data processing is the data description. In order to provide richer automatization capabilities in VirGIS integration system, data used in the system were studied and considered in the Semantic Web context.

The paper is organized as follows: Section 2 provides a brief introduction to the Semantic Web idea, Section 3 gives descriptions of VirGIS system and its data, and Section 4 presents VirGIS data expressed with Semantic Web features.

2 Semantic web

The Semantic Web [1,2] is intended as an extension of today's World Wide Web. It should consist of machine readable, understandable and meaningfully processable data. The basis is addition of data semantics - there will be stored data meaning description together with data themselves. The Semantic Web belongs still to the future; however, there have been made already some features. It is based on standards, which are defined by W3C (WWW Consortium [3]). Semantic Web principles are implemented in layers of web technologies and standards. The layers are figured in Figure 1.

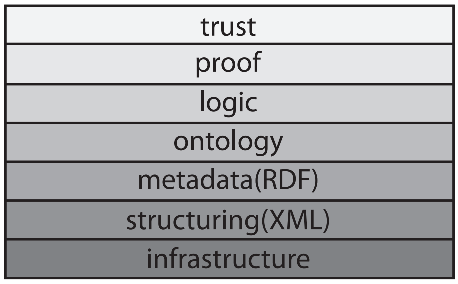

Fig. 1. Semantic web layers

The layer of infrastructure provides a source identification and location. The layer of structuring, the layer of metadata and the ontology layer are essential for describing web sources content. The logic layer enables writing of rules. The layers of proof and trust are things for particular applications. They consider proofs and trust about web information.

2.1 Infrastructure

The Semantic Web should consist of connected sources - it should contain sources and links. Every object should be identified (as on today's web) with identifiers URI (Universal Resource Identifier). The Semantic Web should be decentralized, of course with possibility of missing or incomplete information. It should be able to define source types and links types of course.

2.2 Data description

An important requirement of machine processable information is data structuring. On the web, the main structuring method is using tags, which are parts of text containing information about the role of the text. Nowadays, the metalanguage XML (eXtensible Markup Language) [4] is used for making web document structure. It provides syntax for machine readable data. But only XML is not enough to describe data.

The technique to specify the meaning of information is RDF (Resource Description Framework) [5]. It is basic tool of web sources metadata addition. RDF data model gives an abstract conceptual framework for metadata definition and usage. It uses XML syntax (RDF/XML) for encoding. Additionally, there is also an extension of RDF called RDF Schema [6] that is useful for class definition and class hierarchy description.

An instrument for definition of terms used either in data or in metadata are ontologies [7]. In the context of web technologies, ontology is a file or a document that contain formal definitions of terms and term relations. The Semantic Web technique for definition of ontologies is the OWL (Ontology Web Language) [8] language. Thanks to usage of ontologies, applications can share terms and so it enables application cooperation. Moreover, the Semantic Web idea considers also addition of logic and using inference rules. It brings a possibility to infer and to make conclusions.

2.3 Application operation

The real potential of the Semantic Web would express if people made many programs that would process web sources content and cooperate with other programs. These software agents would be as effective as the web data would be machine understandable and as automated services would be accessible. The Semantic Web should provide a basis for the other technologies.

3 VirGIS

VirGIS [9] is a mediation platform that provides an integrated view of geographic data. The VirGIS system is composed of data sources and a mediator over them. This mediator provides a global virtual view allowing local sources to be accessed as one integrated source.

For querying, a client uses global terms and schema. The mediator rewrites this query, poses it against local data sources, then composes final answer from local answers, and returns the result to the client.

VirGIS is solved as a particular application of data integration. However, some operations have to be done mainly manually. For instance, it is not easy to add a new source. Thus, some improvement is needed. Considering also the fact that the VirGIS system could be accessible via WWW in the future, we were motivated to exploit Semantic Web features. At first, data description was revised and rebuilt.

3.1 VirGIS data

Currently, VirGIS is implemented as an integration system of satellite images. Figure 2 illustrates local and global sources of VirGIS. As local sources are used subsets of schemas drawn from SPOT and IKONOS catalogues and QUICK_LOOK database.

SPOT and IKONOS catalogues provide information about satellites; QUICK_LOOK refers to a sample of small images that give an overview of satellite images supplied in the catalogue. The role of the global source is played by the VIRGIS mediated schema. The VIRGIS schema contains just one entity VIRGIS with following attributes:

SPOT	
Attribute	Type
date_	Date
sun_elev	numeric
satellite	string
sat_id	numeric
key	string
the_geom.	Polygon

IKONOS	
Attribute	Type
date_acqui	Date
sun_el	numeric
satellite	string
sat_id	numeric
key	string
the_geom	Polygon

VIRGIS	
id	string
name	string
satid	string
date	Date
sun_elevation	numeric
url	string
geom	Polygon

QUICK LOOK	
Attribute	Type
key	string
filename	string

Fig. 2. Local and global satellite schemas

- string *id* (a common id for the different region photographed)
- string *name* (the name of the satellite that takes the photo)
- string *satid* (the id for the satellite)
- date *date* (the date when the photo was taken)
- numeric *sun_elevation* (the sun elevation when photo was taken)
- string *url* (the url where the real photo is saved)
- polygon *geom* (the geometry of the region photographed)

VirGIS uses GML [10] as an internal format to represent and manipulate geographic information. GML is a geographic XML-based language; therefore GQuery [11], a geographic XQuery-based language [12], is used for querying.

4 VirGIS data in Semantic web environment

The Semantic Web promises a basis for machine understandable data. It could improve or make easier to automate some operations. Hopefully it could bring something more also in data integration process in VirGIS. There are some data integration areas, which could benefit by better automatization; for example addition of new sources, mapping rules generation and schema evolving. And because the Semantic Web is about standards, we could reuse some tools, which are already made.

An important requirement for machine processable information is data structuring. On the web nowadays, the the language XML is used for making web document structure. But only XML is not enough to describe data. The technique to specify the meaning of information is RDF. Its extension, RDF Schema, is useful for class definition and class hierarchy description. For richer definition of terms, ontologies are used.

In the VirGIS integration system, an XML-based language is used for data representation. If the integration is XML-based, why not bring more and, instead of simple XML, use RDF, which has bigger expressive power. Also XML document primarily not intended for RDF applications could be described using RDF. By observing several guidelines when designing the schema, [13] proposed how to make an XML ”RDF-friendly”. For already existing documents, there is possibility to make some XML-RDF bridge. Of course, it has not to be always simple way.

As with data, the XML and RDF worlds use different formalism for expressing schema. VirGIS expresses local and global schemas with XML Schema. The Semantic Web currently uses languages such as RDFS and OWL. So in the proposed data description, OWL is used to publish sets of terms (called ontologies).

With OWL ontologies, it is possible to share and reuse knowledge. We could exploit other (already defined) ontologies. A large number of organizations have been exploring the use of OWL, with many tools currently available. The Working Group of W3C is maintaining a list of implementations and demonstrations [14].

4.1 The VirGIS Ontology

The aim was a description of VirGIS data - a description of satellite image knowledge in a VirGIS ontology. In ontology re-use, we can consider only some general spatial ontology for basic geometric features. The VirGIS data area itself is not covered with any existing ontology. A new ontology for this purpose is needed.

The proposed VirGIS specified ontology comes out of the data model described above. The main domain concepts and their relationships are depicted in Figure 3 by means of ISA tree.

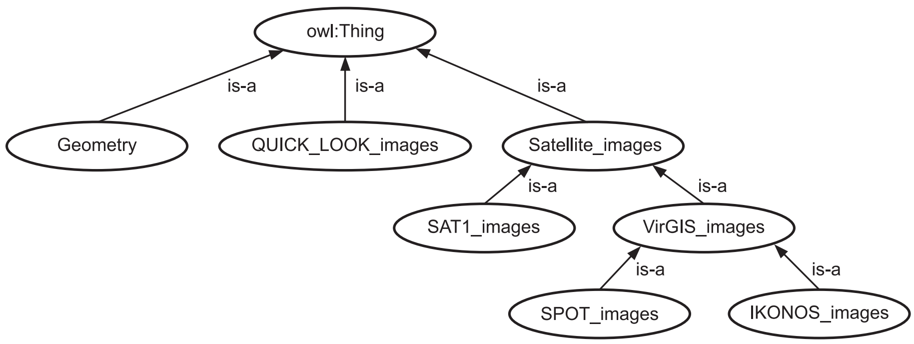

Fig. 3. ISA diagram of the model

Observe that each node corresponds to one concept. `IKONOS_images` and `SPOT_images` refer to local sources; `VirGIS_images` refers to the global mediated source. The fact that every image contained in IKONOS or SPOT database is also contained in VirGIS induces the corresponding concepts relationship that can be understood as set inclusions:

$$IKONOS\_images \subseteq VirGIS\_images,$$
$$SPOT\_images \subseteq VirGIS\_images, \quad (1)$$

Analogical relationship applies to `VirGIS_images` and `Satellite_images` concepts.

Observe that there is an additional class `SAT1_images` in the model. It contains all satellite images not integrated in VirGIS_images. Finally, an inherent feature of the OWL data model is the unique superclass `THING` being the superclass of all other classes.

In OWL, a `owl:Class` construct is used for concept indication and `rdfs:subClassOf` construct for expressing the concept relationships corresponding to set inclusion relations:

Example 1. The OWL expression of the relationship of `SPOT` and `VirGIS` classes

```
<owl:Class rdf:ID="SPOT_images">
   <rdfs:subClassOf rdf:resource="#VirGIS_images" />
</owl:Class>
```

The `rdfs:subClassOf` construct expresses inclusion relationship on both set and conceptual level. Therefore, the above OWL code example implies SPOT_images being conceptually more specific than VirGIS_images.

In OWL, classes are also characterized by means of properties, i.e. attributes of corresponding concepts. Properties definitions are to represent the semantic relationships of the corresponding concepts and their attributes.

Observe that SPOT and IKONOS use semantically equivalent attributes without any common name convention. In addition, VirGIS introduces its own identifiers for respective attributes. For instance, `date_` (SPOT), `date_acqui` (IKONOS) and `date` (VirGIS) represent semantically equivalent attributes for instance. This is solved with mapping of mediation integration in VirGIS. However, it can naturally be expressed on the semantic level, by means of OWL.

With regard to the above discussion and considering the inclusion (1), it follows:

$$(\forall image \in SPOT\_images)(date\_(image, DD/MM/YY) \\ \rightarrow date(image, DD/MM/YY)),$$

which defines the semantic relationship of the binary predicates `date_` and `date`. The relationships between other predicates can be expressed analogically.

In OWL, `rdfs:subPropertyOf` construct is used for expressing such semantic relationships. This relationship is more vague than the relationship of equivalence. However, the relationship of “subPropertyOf” mirrors SPOT_images being conceptually more specific than VirGIS_images.

Example 2. The OWL interpretation of the relationship of the properties `date_` and `date`

```
<owl:DatatypeProperty rdf:about="#date_">
    <rdfs:subPropertyOf rdf:resource="#date" />
</owl:DatatypeProperty>
```

For completeness, there is another additional class. Class `Geometry` class contains geometric elements, designed for geometry type properties description. In case that richer geometry is needed, geometry classes from existing spatial ontologies can be imported. At this time, the presented ontology is suitable for VirGIS data description. It can be enriched in case more capabilities should be needed.

5 Conclusion

Amount of information available is increasing. Manual maintenance and processing is becoming almost impossible; automated processing is needed. The idea of the Semantic Web promises improvement in this area. A key factor is data description. According to this fact and in order to improve data processing in VirGIS integration system, its data were studied. Considering existing Semantic Web features, the data description was then changed and VirGIS ontology was presented.

Acknowledgements

This work was supported by the project 1ET100300419 of the Program Information Society (of the Thematic Program II of the National Research Program of the Czech Republic) "Intelligent Models, Algorithms, Methods and Tools for the Semantic Web Realization", by the project of Czech-French Cooperation Barrande 2004-003-1, 2: "Integration de données sur le Web - applications aux Systémes d'Information Géographique (2003-2005), and by the Institutional Research Plan AV0Z10300504 "Computer Science for the Information Society: Models, Algorithms, Applications".

References

1. M.-R. Koivunen and E. Miller, "W3C Semantic Web Activity", in the proceedings of the *Semantic Web Kick/off Seminar*, Finland, 2001.
2. J. Euzenat. "Research challenges and perspectives of the Semantic Web". Report of the EU-NSF Strategic Research Workshop, Sophia-Antipolis, France, October, 2001.
3. W3C (WWW Consortium), `http://www.w3.org`.
4. Extensible Markup Language (XML), `http://www.w3.org/XML/`.
5. Resource Description Framework (RDF), `http://www.w3.org/RDF/`.

6. RDF Vocabulary Description Language 1.0: RDF Schema, *W3C Recommendation*, http://www.w3.org/TR/2004/REC-rdf-schema-20040210, February, 2004.
7. Y. Ding, D. Fensel, M. Klein, and B. Omelayenko. "The semantic web: yet another hip?". Data & Knowledge Engineering 41 (2002), pp. 205-227.
8. Web Ontology Language (OWL), http://www.w3.org/2004/OWL.
9. O. Boucelma and F.-M. Colonna, "Mediation for Online Geoservices", in Proc. *4th International Workshop Web & Wireless Geographical Information System, W2GIS 2004*, Korea, 2004.
10. Open GIS Consortium Inc., "OpenGIS Geography Markup Language (GML) Implementation Specification", January, 2003, https://portal.opengeospatial.org/files/?artifact_id=7174.
11. O. Boucelma and F.-M. Colonna, " GQuery: a Query Language for GML", in the Proc. *24th Urban Data Management Symposium*, Chioggia-Venice, Italy, October, 2004.
12. "XML Query (XQuery)", http://www.w3.org/XML/Query.
13. B. DuCharme and J. Cowan, "Make Your XML RDF-Friendly", October, 2002, http://www.xml.com/pub/a/2002/10/30/rdf-friendly.html.
14. OWL Implementations. http://www.w3.org/2001/sw/WebOnt/impls.