Ontology approach to integration of geographical data

Zdeňka Linková, Radim Nedbal

Institute of Computer Science, Academy of Sciences of the Czech Republic,
linkova@cs.cas.cz, nedbal@cs.cas.cz

Abstract. A key point in modern automated data processing is metadata semantics representation. Employing Semantic Web existing features - ontologies - is a promising option. Ontologies open a novel approach to knowledge representation.
The paper presents a GIS (Geographic Information System) domain application illustrating ontological approach to data integration and data processing automation in the specific system. This VirGIS system is an integration system that works with spatio-temporal data. We start our study with developing the data representation based on common Semantic Web techniques and build a VirGIS ontology.

1 Introduction

This is an information world. Information is of cardinal importance in many fields of today's human activities. Expansion of World Wide Web has brought better accessibility to information sources. However, at the same time, the big amount of different formats, data heterogeneity, and machine unreadability of this data have caused many problems. Thus automated processing is not an easy task and presents many research challenges today. One of them has initiated the Semantic Web vision [1,2]. This vision aims at data syntax and semantics interconnection. The goal is to reach "machine readability" enabling easier data processing automation.

Sematic Web context is inherent in the VirGIS information system presented. The paper is organized as follows: Section 2 provides a brief introduction to GIS, Section 3 gives descriptions of VirGIS system and its data, and Section 4 presents our ontology approach to VirGIS.

2 Geographical data

Geographical information systems (GIS) [3,4] are generally used to analyze and visualize spatio-temporal information. Originally developed for the creation of thematic maps, GIS support data capture, data storage, and data analysis. The power of GIS comes from the ability to relate different information in a spatial context and to obtain details about this relationship. GIS, therefore, can reveal important new information that leads to better decision making.

Unlike a flat paper map, where what you see is what you get, a GIS can present many layers of different information. These geographic data are thought as layers of information. Each layer represents a particular theme or feature. One theme could be made up of all the roads in an area, another theme could represent all the lakes in the same area. These themes can be laid on top of one another, creating a stack of information about the same geographic area. A GIS combines layers of information about a place to give a better understanding of that place. What layers of information are combined depends on a purpose (e.g. finding the best location for a new store, analyzing environmental damage, etc.). The way data have been stored or filed as layers of information in a GIS makes it easier to perform complex analyses.

The use of GIS can encourage cooperation and communication among organizations. Standardization eases the exchange of digital information among users of different systems. One idea to provide interoperable solutions and applications for geospatial services, data, and applications is to define "simple features" in modelling GIS data. The starting point for modelling of geographic information is a geographic feature. A feature is an abstraction of a real world phenomenon. A geographic feature is a feature associated with a location relative to the Earth. A digital representation of the real world can be thought of as a set of features. The Open Geospatial Consortium [5] Reference Model (ORM) [6] describes a framework for the ongoing work of enhancing and enabling interoperability for technologies involving spatial information and location.

3 GIS data integration

A GIS domain application is VirGIS system [7]. It is a mediation platform that provides an virtual integrated view of geographic data. In general, the main idea in a global virtual view use is a system of components called mediators. Mediators provide an interface of the local data sources. There are also other special components - wrappers, which play the roles of connectors between local source backgrounds and the global one. The principle of integration is to create a nonmaterialized view in each mediator. These views are then used in the query evaluation. Essential are mapping rules that express the correspondence between the global schema and the data source ones. The problem of answering queries is another point of the mediation integration - a user poses a query in terms of a mediated schema, and the data integration system needs to reformulate the query to refer to the data sources.

VirGIS accesses GIS data sources via Web Feature Service (WFS) server and uses WFS interfaces to perform communications with sources. WFSs play the role of wrappers in the mediation system. VirGIS uses GML as an internal format to represent and manipulate geographic information. GML [8] is a geographic XML-based language; therefore GQuery [9], a geographic XQuery-based language, is used for querying. The integration system has only one mediator called GIS Mediator. It is composed of a Mapping module, a Decomposition/Rewrite module, an Execution module and Composition module.

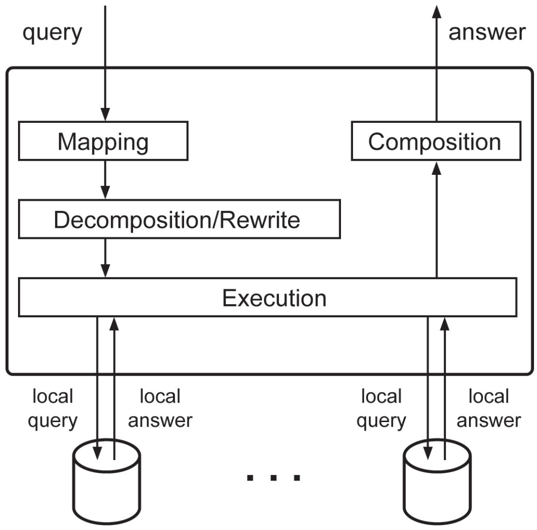

Fig. 1. VirGIS System

The Mapping module uses integrated schema information in order to express user queries in terms of local source schemas. Each mapping rule expresses a correspondence between global schema features and local ones. For the global schema definition, a Local As View (LAV) approach is applied. This approach consists in defining the local sources as a set of views made on the global schema. In current version of VirGIS, there are used simple mapping rules that allow the specification of one-to-one schema transformations under some constraints: aggregations and one-to-many mappings are not considered. The Decomposition/Rewrite module exploits information about source feature types and source capabilities to generate an execution plan. A global GQuery expression is used as a container for collecting and integrating results coming from local data sources. The Execution module processes sub-queries contained in the execution plan and sends them to the appropriate source's WFS. The Composition module treats the final answer to delete duplicities and produces a GML document, which is returned to the user.

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.

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

4 Ontology approach to VirGIS

Our aim is to develop a new ontology for specific GIS area. More generally, we would like to help to develop a new version of the VirGIS integration system that should be Semantic Web based - in order to provide more integration capabilities. 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.

There are many tools and languages [10] that can be employed as means for ontology development. Among available ontology languages, Web Ontology Language (OWL) [11] was chosen, because it is proposed to be an ontology language for the Semantic Web. As an ontology design tool, Protégé System [12] was used.

The proposed VirGIS specified ontology comes out of the VirGIS data model. The VIRGIS schema contains just one entity VIRGIS with attributes string *id*, string *name*, string *satid*, date *date*, numeric *sun_elevation*, string *url*, and polygon *geom*:

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

According to this schema description, the aim was a development of an ontology satisfying the VirGIS data semantics. It had to cover not only the global schema, but also the local ones and relationships among them. The main domain concepts and their relationships are depicted in Figure 3 by means of ISA tree.

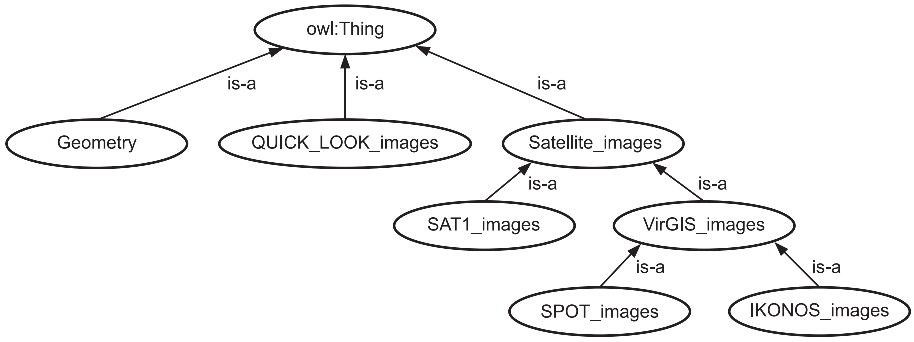

Fig. 3. ISA diagram of the model

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:

$$\begin{aligned} IKONOS\_images \subseteq VirGIS\_images, \\ SPOT\_images \subseteq VirGIS\_images, \end{aligned} \tag{1}$$

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. In addition, `SAT1_images` and `VirGIS_images` are declared as a disjoint classes. Finally, let us remark that an inherent feature of the OWL data model is the all other classes inclusion in the unique, top superclass `THING`.

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

As the amount of information available is increasing manual maintenance and processing is getting almost impossible. Thus the automated processing need is obvious. Nevertheless, the Semantic Web vision promises improvement in this area. A key factor is data representation.

According to this fact, the ontological approach was employed in the geographical data application domain. An ontology instance of real geographical data has been presented. Its influence on data processing effectiveness in VirGIS integration system is studied.

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. J. Euzenat. "Research challenges and perspectives of the Semantic Web". Report of the EU-NSF Strategic Research Workshop, Sophia-Antipolis, France, October, 2001.
2. M.-R. Koivunen and E. Miller, "W3C Semantic Web Activity", in the proceedings of the *Semantic Web Kick/off Seminar*, Finland, 2001.
3. B. Korte George, "The GIS (Geographic Information Systems)", *OnWord Press*, Santa Fe, 1994.
4. Geographic Information Systems. http://erg.usgs.gov/isb/pubs/gis_poster/.
5. Open Geospatial Consortium, Inc. (OGC). http://www.opengeospatial.org/.
6. OpenGIS® Reference Model. http://portal.opengis.org/files/?artifact_id=3836.
7. O. Boucelma and F.-M. Colonna, "Mediation for Online Geoservices", in Proc. *4th International Workshop Web & Wireless Geographical Information System, W2GIS 2004*, Korea, 2004.
8. Open GIS Consortium Inc., "OpenGIS Geography Markup Language (GML) Implementation Specification", January, 2003, https://portal.opengeospatial.org/files/?artifact_id=7174.
9. 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.
10. O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "Methodologies, tools and languages for building ontologies. Where is their meeting point?", *Data & Knowledge Engineering*, vol. 46, pp. 41–64, 2003.
11. Web Ontology Language (OWL), http://www.w3.org/2004/OWL.
12. The Protégé Ontology Editor and Knowledge Acquisition, http://protege.stanford.edu/index.html.