Building Ontology for VirGIS System

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. Ontologies play an important role in a knowledge representation. It involves ontology development as well as ontology re-use. Among various fields, where ontologies can be useful, is the GIS (Geographical Information System) data area. The goal of the research described in this paper is to develop a specific ontology for a given GIS domain. At first, we describe a general methodology and main tools for ontology development. Then a new ontology that covers data used in a VirGIS integration system is presented. The paper describes the VirGIS specified ontology as well as a list of spatio-temporal data ontologies that are available and possible to use for a general data features description.

1 Introduction

The development and usage of geographical information systems (GIS) is based, among other things, on knowledge representation. Ontologies, as specifications of conceptualizations, are possible tools to be employed in this context. Ontologies play an important role in information processing. They enable sharing terms used for information description and thereby they also provide a basis for data sharing, data processing, and, of course, data integration.

Our aim was to build an ontology for a given GIS data domain. It had cover at least data provided by the integration system VirGIS [1].

The paper is organized as follows: Section 2 gives a brief description of GIS; Section 3 provides basic ontology theory and ontology development. Section 4 presents our research description. It introduces VirGIS data and their modelling for ontology building purpose.

2 GIS

Geographical information systems (GIS) [2,3] 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 [4] Reference Model (ORM) [5] describes a framework for the ongoing work of enhancing and enabling interoperability for technologies involving spatial information and location.

3 Ontologies

Ontologies [6,7] were developed in the framework of artificial intelligence (AI) to facilitate knowledge sharing and reuse. Ongoing research on ontologies can be found in the computer science community. The reason of ontologies popularity is that they promise a shared and common understanding of some domain that can be communicated between people and application systems. Ontologies are crucial for knowledge interoperation; sharing the same ontology is a precondition to data sharing and data integration. Ontologies are also central to the Semantic Web [7,8], because they allow applications to agree on the terms and consequently to communicate. They are a key factor for enabling interoperability in the Semantic Web.

The term "ontology" has been used in many ways and across different communities. A popular definition of the term ontology in computer science is: an ontology is a formal, explicit specification of a conceptualization. A conceptualization refers to an abstract model of some phenomenon in the world. However, a conceptualization is never universally valid. Ontologies have been set out to overcome the problem of implicit and hidden knowledge by making the conceptualization explicit. An ontology may take a variety of forms, but it will necessarily include a vocabulary of terms and some specification of their meaning.

Ontologies may help also in GIS data sharing and processing. The need to share geographic information is evident. Today, there is a huge amount of data gathered about the Earth, computers throughout the world are connected, and the use of GIS has become widespread. The support and use of multiple ontolo-

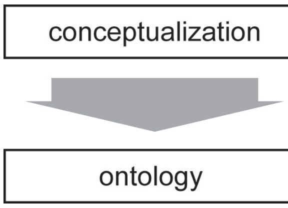

Fig. 1. Conceptualization and ontology

gies should be a basic feature of systems that should be able to solve semantic heterogeneity to make use of the amount of information available.

3.1 Ontology development

Ontologies aim at modelling and structuring domain knowledge. The purpose is to provide understandable domain description, which may be used and shared across applications and groups of people. It requires methodologies that cover all aspects [9]. Therefore an ontology development follows a cycle [7] containing several phases.

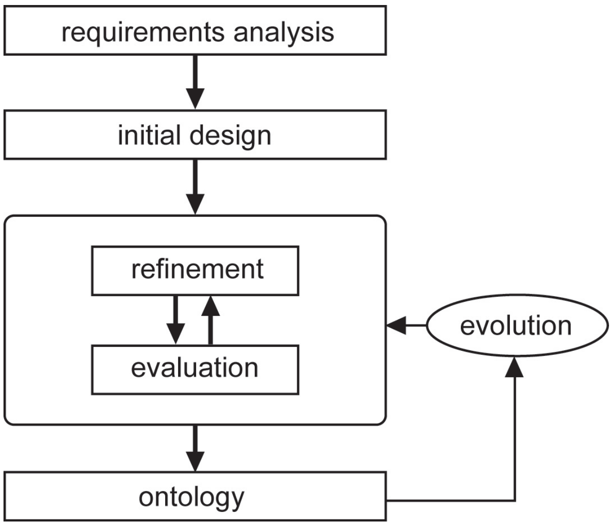

Fig. 2. Ontology lifecycle

- The ontology lifecycle starts with a requirements analysis. In this step, it is important to consider the breadth of the ontology and richness of the modelled domain. A key issue is the balance between specificity and reusability.

- Based on the requirements specification, the initial ontology is designed. Thorough analysis of various kinds of knowledge sources (free texts, semistructured sources, and structured sources) is important. Here ontology learning, i.e. the combination of linguistic analysis, information extraction, statistical techniques and machine learning is a promising area of research. The output of the initial design step is some preliminary conceptual structure.
- In the next step, the basic ontology design is refined and evaluated. To support efficient and effective refinement, tools for restructuring and enriching data are needed. It is possible to exploit already existing conceptual sources, like thesauri, database schemata, other ontologies etc. Created conceptual structure needs to be evaluated with respect to its requirements and eventually more specified.
- The ontology lifecycle does not end with refinement and evaluation steps. In real world, things are changing, and so should do ontologies. To handle the evolution and maintenance of ontologies, one needs to explore and to formalize the kinds of relationships that may rise between different ontology versions.

3.2 Ontology languages and tools

Contemporary ontology languages are based on the XML syntax. It is a consequence of the fact that XML (eXtensible Markup Language) [10] has become a standard language for information exchange on the Web. Also the common RDF (Resource Description Framework) [11] syntax is based on the XML. RDF was developed by the W3C (the World Wide Web Consortium) [12] as a framework to describe Web resources.

Its extension, RDF Schema [13], is RDF's vocabulary description language. RDFS provides mechanisms for describing groups of related resources and the relationships between these resources. It allows the representation of concepts, concept taxonomies and binary relations.

However it is not very expressive. For more exact description of knowledge, a richer language is needed. Therefore three more languages have been developed as alternative to RDF(S): OIL (Ontology Interchange Language) [14], DAML+OIL [15], and OWL. OWL (Web Ontology Language) [16] is a product of W3C and is presented as an ontology language for the Semantic Web. It allows representing not only concepts, taxonomies, binary relations, but also cardinalities, richer type definitions and other characteristics.

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 [17]. Most of the systems currently using DAML, OIL and DAML+OIL are now migrating to OWL. In addition, a number of ontology language tools, such as the widely used Protégé system [18], now provide OWL support.

It is reasonable to assume that ontologies could be available on the market. As ontology development technology evolves, the benefits of ontology use will outweigh the costs of developing them. With the success of this approach, large-scale repositories of ontologies will be available in diverse disciplines. Also a commercial production is possible. However, the available quantity of ontological knowledge is modest and their quality, too. Some types of objects have been the objects of ontology study, some objects have received little attention.

An option is to use an ontology library containing specialized ontologies of domain and tasks. There is a large number of ontologies available on the Web. There is a DAML ontology library [19], which contains about 280 examples written in OWL or DAML+OIL (a converter from DAML+OIL to OWL is also available on the Web). The library organizes hundreds of ontologies in a variety of different ways (keyword, organization, submission date, etc.). In addition, several large ontologies have been released in OWL. And as in other research areas, there have been also some projects of ontology development in GIS data field (e.g. [20,21,22,23]).

4 Building ontologies for GIS

As an ontology design tool, Protégé System [18] was used. Protégé is an integrated software tool used by system developers and domain experts to develop ontologies and knowledge-based systems. Protégé has been developed by the Stanford Medical Informatics (SMI) at Stanford University. It is an open source, standalone application with an extensible architecture. It holds a library of plugins that add more functionality to the environment. Protégé's OWL Plug-in now provides support for editing Semantic Web ontologies. There is also a list of the currently made ontologies on the Protégé Ontologies Library page [24]. It is a small but hopefully growing selection of existing OWL ontologies that one can use.

As a part of the research, OWL ontologies on the Internet were explored. Some of them are suited for the geographical information system. They fulfill geographic information standards (ISO, OpenGIS®Consortium - OGC, or standard by Federal Geographic Data Committee - FGDC) [22]. The list is given in Figure 3.

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 [1] that should be Semantic Web based - in order to provide more integration capabilities. For more general data description, our first step was development of an ontology covering at least data provided by this system.

4.1 VirGIS data

VirGIS 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, which is called a GIS mediator, provides a global virtual view allowing local sources to be accessed as one integrated source.

ISO/CD TS 19103	Ontology for Conceptual Schema Language
ISO 19107:2003	Ontology for Geographic Information - Spatial Schema
ISO 19108:2002	Ontology for Geographic Information - Temporal Schema
ISO/FDIS 19109	Ontology for Geographic Information - Rules for Application Schema
ISO/FDIS 19110	Ontology for Geographic Information - Methodology for Feature Cataloguing
ISO 19111:2003	Ontology for Geographic Information - Spatial Referencing by Coordinates
ISO 19112:2003	Ontology for Geographic Information - Spatial Referencing by Geographic Identifier
ISO 19115:2003	Ontology for Geographic Information - Metadata
ISO 19115:2003	Ontology for Geographic Information - Metadata Application
FGDC	Ontology for Content Standard for Digital Geospatial Metadata
OGC	Ontology for Topic-2: Spatial Referencing by Coordinates
OGC	Ontology for Geography Markup Language (GML3.0)

Fig. 3. List of OWL ontologies based on norms

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

Currently, VirGIS is implemented as an integration system of satellite images. Figure 4 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	
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. 4. Local and global satellite schemas

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:

- 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)

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.

4.2 The VirGIS Ontology

The aim was 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 GIS 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 5 by means of ISA tree.

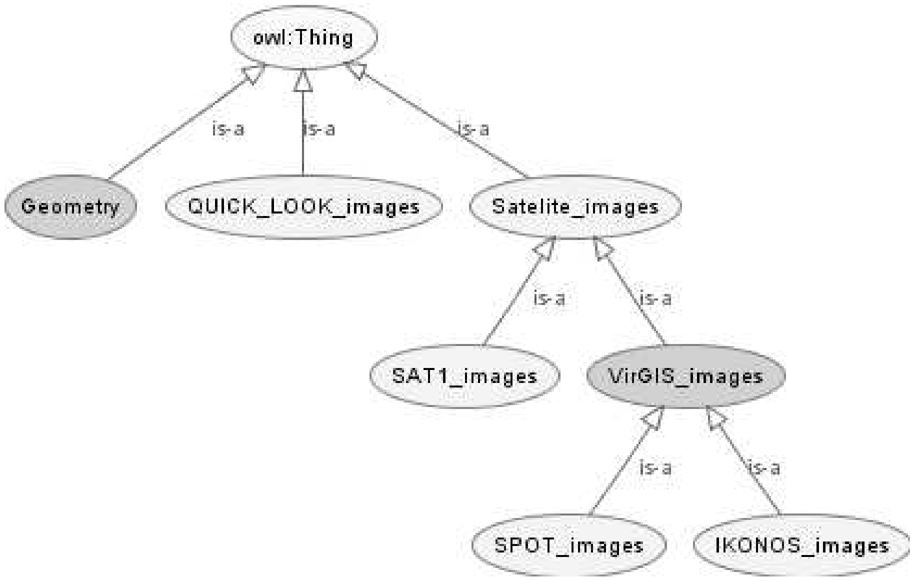

Fig. 5. ISA diagram of the model

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:

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

Analogical relationship applies to `VirGIS_images` and `Satelite_images` concepts. Observe that there is an additional class `SAT1_images` in the model. It contains 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.
```
<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. `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:

Example 2.
```
<owl:DatatypeProperty rdf:about="#date_">
    <rdfs:subPropertyOf rdf:resource="#date" />
</owl:DatatypeProperty>
```

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.

For completeness, there is an additional class in the model. `Geometry` class contains geometric elements, designed for geometry type properties description. In case that richer geometry is needed, geometry classes from existing spatial ontologies listed in Figure 3 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

Ontologies are crucial in data description. According to this fact, a new ontology in GIS data area was developed. Particularly, this ontology describes sources and data in the VirGIS integration system. As other ontologies, also this one should be evolved in order to follow the evolution of things it describes. In this case, it can for instance mean ontology enrichment along with adding new sources to VirGIS. Although ontologies are very powerful tools in data processing, there is still a lack of available and suitable ontologies in many areas. We believe that our research and development can contribute to increase number of usable ontologies and can help in VirGIS data integration task.

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. O. Boucelma and F.-M. Colonna. "Mediation for Online Geoservices". in Proc. 4th International Workshop Web & Wireless Geographical Information System, W2GIS 2004, Korea, November, 2004.
2. Your Internet Guide to Geographic Information Systems. `http://www.gis.com`.
3. Geographic Information Systems. `http://erg.usgs.gov/isb/pubs/gis_poster/`.
4. Open Geospatial Consortium, Inc. (OGC). `http://www.opengeospatial.org/`.
5. OpenGIS® Reference Model.
`http://portal.opengis.org/files/?artifact_id=3836`.
6. Y. Ding, D. Fensel, M. Klein, and B. Omelayenko. "The semantic web: yet another hip?". Data & Knowledge Engineering 41 (2002), pp. 205-227.
7. J. Euzenat. "Research challenges and perspectives of the Semantic Web". Report of the EU-NSF Strategic Research Workshop, Sophia-Antipolis, France, October, 2001.

8. M.-R. Koivunen and E. Miller. "W3C Semantic Web Activity". in the proceedings of the Semantic Web Kick/off Seminar, Finland (2001).
9. O. Corcho, M. Fernández-López, and A. Gómez-Pérez. "Methodologies, tools and languages for building ontologies. Where is their meeting point?". Data & Knowledge Engineering 46 (2003), pp. 41–64.
10. Extensible Markup Language (XML). `http://www.w3.org/XML/`.
11. Resource Description Framework (RDF). `http://www.w3.org/RDF/`.
12. W3C (WWW Consortium). `http://www.w3.org`.
13. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, `http://www.w3.org/TR/2004/REC-rdf-schema-20040210`, February, 2004.
14. OIL. `http://www.ontoknowledge.org/oil`.
15. About the DAML Language. `http://www.daml.org/about.html`.
16. Web Ontology Language (OWL). `http://www.w3.org/2004/OWL`.
17. OWL Implementations. `http://www.w3.org/2001/sw/WebOnt/impls`.
18. The Protégé Ontology Editor and Knowledge Acquisition. `http://protege.stanford.edu/index.html`.
19. DAML Ontology Library. `http://www.daml.org/ontologies/`.
20. U. Visser, H. Stuckenschmidt, G. Schuster, and T. Vögele. "Ontologies for geographic information processing". Computers & Geosciences 28 (2002), pp. 103-117.
21. F. T. Fonseca, M. J. Egenhofer, C. A. Davis Jr., and K. A. V. Borges. "Ontologies and knowledge sharing in urban GIS". Computers, Environment and Urban Systems 24 (2000), pp. 251-271.
22. List of OWL Ontologies based on Norms. `http://loki.cae.drexel.edu/w̃bs/ontology/list.htm`.
23. N. Lammari, E. Métais. "Building and maintaining ontologies: a set of algorithms". Data & Knowledge Engineering 48 (2004), pp. 155–176.
24. Protégé Ontologies Library. `http://protege.stanford.edu/ontologies/ontologies.html`.