Building Ontologies for GIS

Linková, Zdeňka
2005

Institute of Computer Science
Academy of Sciences of the Czech Republic

Building Ontologies for GIS

Zdeňka Linková, Radim Nedbal, Martin Řimnáč

Technical report No. 932

March 2005

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 053 630, fax: +420 286 585 789,
e-mail:{linkova,nedbal,rimnac}@cs.cas.cz,

Institute of Computer Science
Academy of Sciences of the Czech Republic

Building Ontologies for GIS

Zdeňka Linková, Radim Nedbal, Martin Řimnáč

Technical report No. 932

March 2005

Abstract:

Knowledge representation in geographic information systems (GIS) and associated data processing presents many challenges for researchers. To use ontologies as knowledge representation belongs to the most topical problems to solve. This involves ontology development as well as ontology re-usage. The goal of the research described in this paper is to develop a specific ontology for a given GIS area.

Keywords:
Ontology building, GIS

Contents

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 is to build an ontology for a given GIS data domain. It should 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, ontology development, and ontology re-usage. Section 4 presents our research description, and introduces VirGIS data and their modeling for ontology building purpose.

2. GIS

GIS [2], [3] is computer software that links geographic information (where things are) with descriptive information (what things are). Geographical information systems 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.

A GIS can also convert existing digital information, which may not yet be in a map form, into forms it can recognize and use. For example, digital satellite images can be analyzed to produce a map of digital information about land use and land cover. Because digital data are collected and stored in different ways, the two data sources may not be entirely compatible. Therefore, a GIS must be able to convert data from one structure to another. Satellite image data that have been interpreted by a computer to produce a land use map can be "read into" the GIS in raster format. Raster data files consist of rows of uniform cells coded according to data values. Raster files can be manipulated quickly by the computer, but they are often less detailed and may be less visually appealing than vector data files, which can approximate the appearance of more traditional hand-drafted maps. Vector digital data are captured as points, lines (a series of point coordinates), or areas (shapes bounded by lines). A GIS can be used to convert a satellite image map to a vector structure by generating lines around all cells with the same classification, while determining the spatial relationships of the cell, such as adjacency or inclusion.

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 you combine 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 a 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 modeling GIS data. The starting point for modeling of geographic information is the 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, in such areas as computational linguistics or database theory. It covers fields ranging from knowledge engineering, information integration, information retrieval, and object-oriented analysis to such applications as medicine, mechanical engineering, and GIS. 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. Philosophers and software engineers have different perspectives about it. The word ontology was taken from Philosophy, where it means a systematic explanation of being. Since this term became relevant also in computer science field, where many definitions of ontologies in this context have been proposed. Ontologies are mostly considered as an explanation of some shared vocabulary or conceptualization of specific subject matter.
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. In general, each person has an individual view of the world. Terms from natural language used to communicate have a common basis of understanding and can be assumed to be a shared vocabulary. This common understanding relies on the idea of how the world is organized. This idea is often called a conceptualization of 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.

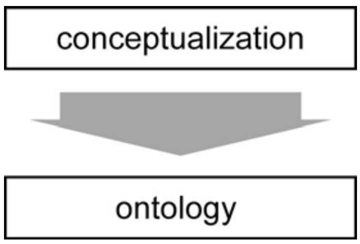

Figure 1. Conceptualization and ontology

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. Although spatial information systems have been characterized also as an integration tool, GIS interoperability is far from being fully operational. The support and use of multiple ontologies should be a basic feature of systems that should be able to solve semantic heterogeneity to make use of the amount of information available.

Ontology development

Ontologies aim at modeling 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, ranging from the requirements analysis and initial ontology design to conceptual refinement, evaluation and evolution.

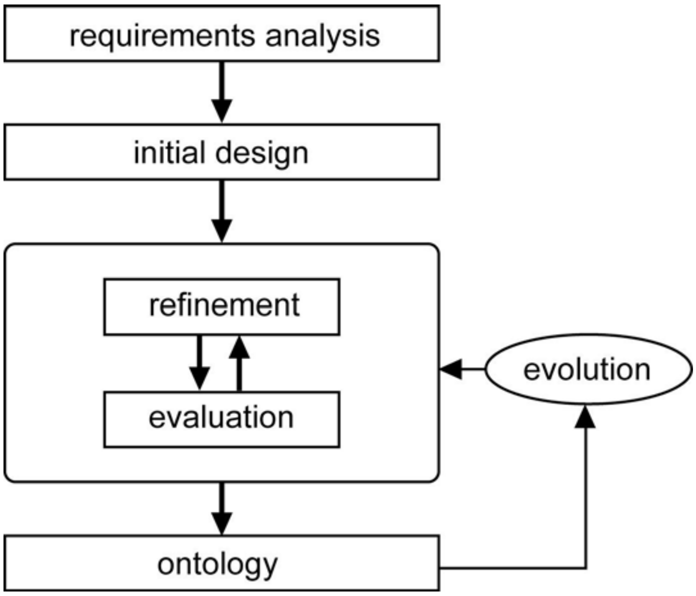

Figure 2. Ontology lifecycle

- The engineering of ontologies starts with a requirements analysis. In this step, it is important to consider aspects like coverage (the breadth of the ontology), richness (how detailed the domain should be modeled), and cognitive adequacy (what are structures and concepts that fit to the needs). A key issue is also the balance between specificity and reusability.
- Based on the requirements specification, the initial ontology design may be approached. 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 very challenging, yet 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 evaluation can be performed in a neutral way (in abstracto) or in a particular application.
- 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. Significant is also indication that some ontology parts may have become outdated and this information must be distributed.

Ontology languages and tools

Today’s 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, DAML+OIL and OWL. OIL (Ontology Interchange Language) [14] was developed as a result of effort to unify modeling bases, web languages, and formal semantics into one language. 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], and [23]).

4. Building ontologies for GIS

After preliminary studies, we have started with the installation of Protégé 2000 System [18]. Protégé-2000 is an integrated software tool used by system developers and domain experts to develop ontologies and knowledge-based systems. Protégé-2000 has been developed by the Stanford Medical Informatics (SMI) at Stanford University, and is the latest version of the Protégé line of tools. 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.

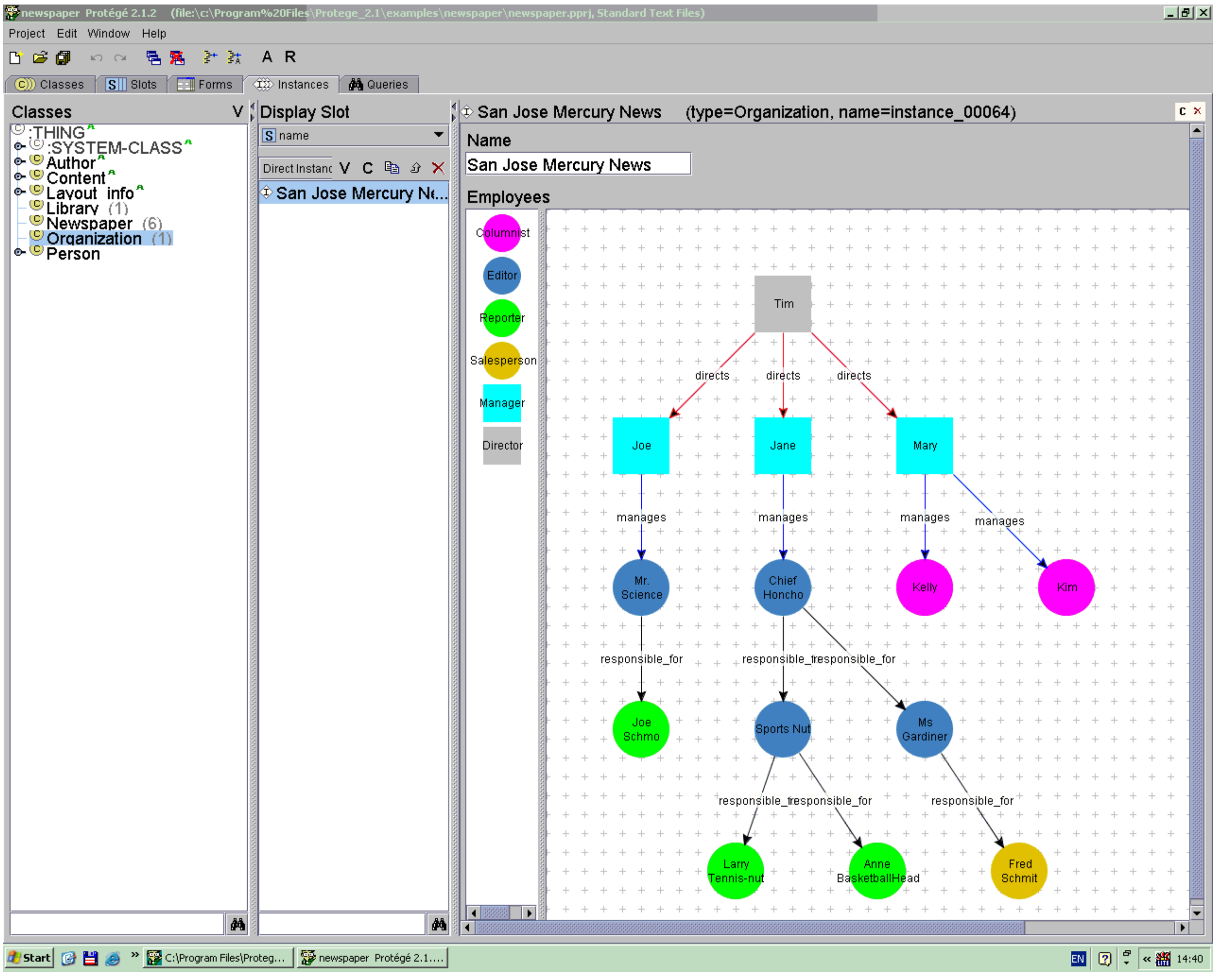

Figure 3. Protégé-2000

The Protégé-2000 tool accesses all of parts through a uniform GUI (graphical user interface) whose top-level consists of overlapping tabs for compact presentation of the parts and for convenient co-editing between them. This tabbed top-level design permits the modeling of an ontology of classes describing a particular subject, the creation of a knowledge-acquisition tool for collecting knowledge, the entering of specific instances of data and creation of a knowledge base, and the execution of applications.

The knowledge-acquisition tool is designed to be domain-specific, allowing domain experts to easily and naturally enter their knowledge of the area. The resulting knowledge base can then be used with a problem-solving method to answer questions and solve problems regarding the domain. Finally, an application is the end product created when the knowledge base is used in solving an end-user problem employing appropriate problem-solving, expert-system, or decision-support methods.

As a part of our research, we have explored OWL ontologies on the Inernet. 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 4.

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)

Figure 4. List of OWL ontologies based on norms

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 ontology based. Therefore, the developed ontology should at least cover data provided by this system.
VirGIS is a mediation platform that provides an integrated view of geographic data. The existing VirGIS prototype integrates satellite images sources. The global schema provided 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)

We started with building ontology for this global schema. We aim at description of satellite image knowledge in an ontology. We could describe the entity from VirGIS global schema as one class for satellite image with properties corresponding to mentioned schema attributes. We want to exploit hierarchical structure capabilities, and we do not aim at specifying just one particular class. Some properties could be inherited from class describing more general image or ranking in satellite ontology class. In general, the complexity of the geographic objects asks for the combination of multiple ontologies. To build a geographic ontology, it is necessary to combine, for instance, a spatial ontology with a geometric ontology and a spatial reference system ontology. The same situation is in our case.
We decided:

- to use geometric classes from an existing spatial ontology (e.g. for the Polygon attribute *geom* description)
- to use an existing temporal ontology (e.g. for the date attribute date description)
- to add our domain specified ontology

However, the above mentioned global schema provides just a simple view over richer capabilities of data sources. We have concluded that the expressiveness of the ontology describing this global schema is not sufficient. Different satellites have a different set of attributes; some attributes with different name have the same meaning. We should deal with this fact and use ontology capabilities to express property equivalency. According to this approach, we are currently dealing with original data sources in order to make a richer conceptualization of satellite image domain and consequently to make its otology.

5. Conclusion

Ontologies are very powerful tool in data description. They provide basis for machine processible data, enable data sharing, data integration, and application communication. Aware of these facts, we aim at building a GIS ontology.

Our research consists of exploration of ontology development and usage and building a particular ontology. We started with simple ontology and nowadays we are about to build richer one. The goal is to provide ontology that could be useful across many GIS applications and that could enable (among others) better data integration.

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. Ed. George Percivall, September, 2003. 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, February, 2004. http://www.w3.org/TR/2004/REC-rdf-schema-20040210.

[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/~wbs/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.