Geo-semantics and Ontology
Extended abstract

Thomas Bittner
Institute for Formal Ontology and Medical Information Science (IFOMIS)
University of Leipzig
and
Stephan Winter
Department for Geomatics
The University of Melbourne

Introduction

One of the major problems facing systems for Architecture Engineering and Construction (AEC) and Geographic Information Systems (GIS) applications today is the lack of interoperability among the various systems. In order to integrate software applications substantial difficulties can arise in translating information from one application to the other. In this paper we focus on *semantic* difficulties that arise in this process. Applications may use different terminologies and representations of the same domain. Even when applications use the same terminology, they often associate different semantics with the terms. This prevents from exchanging information among applications. What is needed is some way of explicitly specifying the terminology of the application in an unambiguous fashion. Ontologies can provide such means and it will be the task of this paper to explain what ontologies are and how they can be used to facilitate interoperability between AEC and GI systems.

What is an ontology?

In first approximation an ontology can be considered as a specification of a conceptualization [11]. Conceptualizations represent ways in which we humans understand the world. In GIS for example we use concepts like *parcel*, *highway*, *lake*, etc. In AEC we use concepts like *building*, *room*, *garden*, *backyard*, etc. A specification then is some abstract description of those concepts.

Uschold [20] acknowledges that there is a wide variety of ways in which such a specification can be achieved but stresses that ``necessarily it will include a *vocabulary of terms* and some *specification of their meaning*''. Guarino [12] then adds that an ontology is an engineering artifact which is constituted by a specific vocabulary used to describe a certain part of reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words.

Ontologies can be distinguished with respect to the degree of rigor in which they are stated. Ideally they are specified in form of a logical theory. Here, as Guarino [12] puts it, the vocabulary words appear as unary and binary predicate names, respectively called concepts and relations. Logical axioms are then added in order to express other relationships between concepts and explicate their intended meaning. Within such a formal and rigorous framework logical deduction can be used in order to check the consistency of such ontologies and to compute consequences of the assumptions that have been made.

Often ontologies are specified in non-rigorous languages, for example in natural language as in the various ISO standards or in semi-formal languages such as UML. With the growth of the Semantic Web the specification of ontologies in Description Logics [1] (a specific form of formal logic that can be run on a computer) have become popular. This is because ontologies specified using a description logic can be `understood' by a computer program without human assistance.

Purpose and use of ontologies

Ontologies are used to improve communication either between humans or computers. More specifically, Jasper and Uschold [16] identify three major areas of use: (i) to assist in communication between human beings, (ii) to achieve interoperability (communication) among software systems, and (iii) to improve the design and the quality of software systems. In this paper we focus on (i) and (ii).

When ontologies are used to facilitate the communication between humans (i), for example in order to agree on data standards or to design an interface between different software systems, then an ontology stated informally in a non-rigorous way is often sufficient. This is because here ontologies are interpreted by human beings and the disambiguation of terminology is their main purpose.

There are at least two different ontology-based classes of solutions to the problem of enabling different software applications to communicate (ii): One is the standardization approach in which all applications *share a common ontology*. The shared ontology is encoded into the standard and all applications which adhere to the standard use the same terminology in an unambiguous fashion.

In the second class of solutions an ontology is used as an interlingua or *reference ontology*. In such a framework every application has its own specific ontology and provides transformations of their vocabulary into the vocabulary of the reference ontology and vice versa. Applications with different internal ontologies then can interoperate by first transforming the statements (data) expressed in the vocabulary of the source ontology to statements (data) expressed in the vocabulary of the reference ontology, and from here to statements (data) expressed in terms of the vocabulary of the ontology of the target application. Hereby the transformation is performed in such a way that the meaning of the statements (data) is preserved.

When using an ontology as an interlingua or reference ontology, two basic scenarios of generating the translations can be distinguished. Firstly, translators from/to the different software applications can be written by humans. Here the underlying ontology is interpreted by the human beings writing the translators. This is the state of the art today. In the second scenario the computer programs themselves are able to interpret the reference ontology and perform the translation and communication process. This is the core of the dream of the Semantic Web [3,8].

It has been pointed out in the literature that the standard-based approach is useful in restricted domains and relatively homogeneous environments while the use of reference ontologies is more suitable in non-restricted domains and heterogeneous environments [6].

Data standards and reference ontologies for AEC and GIS

To provide the foundations for systems that overcome the historic distinction between AEC and GIS and to facilitate interoperability between AEC and GI systems both ontology-based solutions can be exploited: the standard-based as well as the one based on reference ontologies.

We argue that standardization will be successful in cases where both, AEC and GIS, share common ground that can be made explicit in a shared common ontology. In our opinion this will result in a large degree of standardization of the *spatial* component of AEC and GIS systems. Here already today data standards are applied quite successfully. Consider for example the body of standards that has been worked out within the OpenGIS community [7,2], or the norms set up by ISO [15].

In this extended abstract we shall omit the discussion of the standard-based approach and focus on the use of reference ontologies in the context of specifying the semantics of the *attribute* component in AEC and GIS data. However we stress that besides the commonalities shared by AEC and GIS due the spatial nature of their data, there are, however, also important differences. These include differences regarding to scale, the mainly three-dimensional character of CAD representations, the mainly two-dimensional character of GIS representations. These are areas in which data standards shared by AEC and GIS applications need to go beyond standards that exist today. For example, the Simple Features in the OpenGIS standard [2] are still two-dimensional.

Now consider the use of reference ontologies in facilitating AEC and GIS interoperability. In this context it is important to recognize that, strictly speaking, the spatial components of AEC and GIS `only' specify the spatial location of spatio-temporal entities. From the ontological perspective the non-spatial (or thematic) attribute data is much more complex.[1] This complexity comes from the need to specify the nature of the

[1] Let us assure that we do not underestimate the complexity of the representation of the spatio-temporal extension. In particular the different nature of large scale and small scale objects and the phenomenon of vagueness and indeterminacy of location need further research [5]. However, specifying the spatial (and/or temporal) location of things is only one aspect of many.

entities themselves and the heterogeneity of the different kinds of things in the target domains of AEC and GIS.

In order to achieve interoperability at the level of attribute data we need to provide reference ontologies that are general enough to specify the conceptualization of entities that range from concepts for small-scale objects like a cups or tables to concepts for large-scale human artifacts such as buildings, cities, highway systems, or political subdivisions, to concepts for large scale natural phenomenas such as rivers, lakes, wetlands, environments, or ecosystems.

These different kinds of concepts belong to different *domain ontologies*, e.g., ontologies of ecosystems, transportation ontologies, or ontologies of real estate. Domain ontologies describe the basic concepts of a particular domain. Ontologies sufficient for the purpose of facilitating interoperability of attribute data of AEC and GI systems need to provide an account of notions that are fundamental in any of those domains.

Despite all the distinctions and the variety of different notions all domain ontologies use the same *top level* concepts and relations. Examples of those top-level notions include: categories like endurants and perdurants [14], which refer to different modes of existence in time; mereological (part-of) relations [19,13], topological (connectedness) relations [21], and dependence relations; the notions of granularity and context [4].

Ontologies at this level of generality are called *top-level* ontologies [17]. They provide the basis upon which domain ontologies are built. In order to provide transformations between ontologies of domains in which AEC systems are traditionally used and ontologies of domains in which GIS are traditionally used it is critical to incorporate top-level ontologies into the core of the reference ontology.

Top-level ontologies

Currently there are multiple top-level ontologies under development. Examples are DOLCE [17] and Basic Formal Ontology (BFO) [10]. Here we will focus on BFO but we could have chosen DOLCE equally well.

The most basic categorical distinction between entities at the top-level relates to different modes of persistence through time. Two categories of persistent entities can be distinguished: endurants and perdurants (processes). Endurants are wholly present (i.e., all their proper parts are present) at any time at which they exist. For example, you (an endurant) are wholly present in the moment you are reading this. No part of you is missing. Endurants can change and yet remain the same. For example all the cells in your body are replaced over a period of 10 years nevertheless you are the same person today you were 10 years ago.

Perdurants, on the other hand, are extended in time in virtue of possessing different temporal parts at different times. In opposition to endurants they are only partially present at any time at which they exist - they evolve over time. For example, at this moment only a (tiny) part of your life (a perdurant) is present. Larger parts of your life - such as your childhood - are not present at this moment.

Understanding the distinction between endurants and perdurants is important when time and change over time is incorporated into information systems [18,10]. The notions of process and change are critical in domains in which GIS have been used traditionally, for example in hydrology and in environmental science [9]. To overcome the historical distinction between AEC and GI systems both need to take into account the notions of process and change. Incorporating these notions into reference ontologies that provide a bridge between the two is the first step toward applications that have the strengths of both kinds of systems.

Endurants are divided into two major categories: independent endurants such as cups, buildings, bridges, and highway systems, and dependent endurants such as qualities, roles, states or functions. Here we focus on the former. BFO distinguishes the following kinds of independent endurants: substances, fiat parts of substances, aggregates of substances, and boundaries of substances.

- Substances are maximally connected entities, i.e., they have connected bona fide boundaries, i.e., boundaries which correspond to discontinuities in the underlying reality.
- Neither your nose nor your arm are substances. Both are fiat parts of you, i.e., (at least partly) bound by boundaries that do not correspond to discontinuities in the underlying reality but to a human definition on a continuum. Similarly, mountains are fiat parts of the planet Earth, or land parcels are fiat parts of the surface of the earth.
- Aggregates of substances are not substances either. Examples of aggregates are: your family, the heating facilities in a given building, the water supply facilities in a town, etc.

Historically, CAD - and thus AEC systems - are good at modeling aggregates, while fiat subdivisions such as land parcels were modeled primarily in GIS. To overcome this distinction it is important to incorporate the concepts of substance, fiat part, and aggregate into both systems. Top level ontologies like BFO give a formal account of relationships between substances, their fiat parts, and the aggregates they form. Again, incorporating these notions into reference ontologies that provide the bridge between AEC and GI systems is the first step toward systems that have the strengths of both.

Summary

In this paper we discussed how ontologies can be used in order to overcome the historic distinction between AEC and GI systems and to facilitate the semantic interoperability of the two kinds of systems. We started with a discussion of what ontologies are, what their purpose is, and how they are used. We distinguished two major strategies of applying ontologies in order to facilitate interoperability: the use of data-standards and the use of reference ontologies. The former strategy is based on a shared ontology which is encoded into a standard and all applications which adhere to the standard are interoperable by using the same terminology in an unambiguous fashion. In the second strategy a reference ontology is used as an interlingua which provides means of transformation between different application ontologies.

In the context of AEC and GIS we argued that the standard-based strategy is sufficiently powerful in order to facilitate the interoperability of the spatial component of AEC and GIS data. We also argued that in order to achieve interoperability at the level of the attribute data the more powerful and more flexible strategy of using reference ontologies is needed. In particular we argued that, due to the heterogeneous character of the domain ontologies which describe the attribute data in the domains in which AEC and GI systems are used, top-level ontologies need to be a core component of the applied reference ontologies.

Top-level ontologies describe notions that are so general that they are common to all the different domain ontologies. For this reason they are of particular importance for the design or reference ontologies that are used to facilitate interoperability between AEC and GI systems. As an example we discussed some core notions of a specific top-level ontology, BFO.

Bibliography

1
F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, editors.
The Description Logic Handbook.
Cambridge University Press, 2002.

2
D. Beddoe, P. Cotton, R. Uleman, S. Johnson, and J. R. Herring.
OpenGISR simple features specification for SQL revision 1.1.
OpenGISR Implementation Specification OpenGIS Project Document 99-049, Open GIS Consortium, Inc., 1999.

3
T. Berners-Lee, J. Hendler, and O. Lassila.
The Semantic Web.
Scientific American, May 2001.

4
T. Bittner and B. Smith.
A theory of granular partitions.
In M. Duckham, M. F. Goodchild, and M. F. Worboys, editors, *Foundations of Geographic Information Science*, pages 117-151. London: Taylor & Francis, 2003.

5
P. Burrough and A. U. Frank, editors.
Geographic Objects with Indeterminate Boundaries.
GISDATA Series II. Taylor and Francis, London, 1995.

6
M. Ciocoiu, M. Gruninger, and D.S. Nau.
Ontologies for integrating engineering applications.
Journal of Computing and Information Science in Engineering, 2000.
7
S. Cox, P. Daisey, R. Lake, C. Portele, and A. Whiteside.
OpenGISR geography markup language (gml) implementation specification.
OpenGISR Implementation Specification OpenGISR project document OGC 02-023r4, Open GIS Consortium, Inc., 2003.
8
M. Egenhofer.
Toward the Semantic Geospatial Web.
In *Tenth ACM International Symposium on Advances in Geographic Information Systems*, pages 1-4, McLean, Virginia, 2002. ACM Press.
9
C.C. Feng, T. Bittner, and D.M. Flewelling.
Modeling process with endurance and perdurance: The surface hydrology case.
Technical report, Department of Geography, State University of New York at Buffalo, 2004.
10
P. Grenon and B. Smith.
SNAP and SPAN: Prolegomenon to geodynamic ontology.
Spatial Cognition and Computation, 2004.
11
T. Gruber.
A translation approach to portable ontology specification.
Knowledge Acquisition, pages 199-220, 1993.
12
N. Guarino.
Formal ontology and information systems.
In Nicola Guarino, editor, *Formal Ontology and Information Systems, (FOIS'98)*. IOS Press, 1998.
13
N. Guarino, S. Pribbenow, and L. Vieu.
Modeling parts and wholes.
Data & Knowledge Engineering, 20(3):257-258, 1996.
14
K. Hawley.
How things persist.
Oxford : Clarendon Press, 2001.
15
ISO.
Geographic information - spatial schema.
International Standard ISO 19107, ISO, 2003.
16

R. Jasper and M. Uschold.
A framework for understanding and classifying ontology applications.
In *Proceedings of the IJCAI99 Workshop on Ontologies and Problem-Solving Methods(KRR5)*, 1999.

17

M. Masolo, S. Borgo, A. Gangemini, N. Guarino, A. Oltramari, and A. Oltramari.
Wonderweb deliverable D18 - ontology library (final).
Technical report, ISTC-CNR, 2004.

18

F. Reitsma and T. Bittner.
Process, hierarchy, and scale.
In W. Kuhn., M. Worboys, and S. Timpf, editors, *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science. International Conference COSIT'03*, Lecture Notes in Computer Science. Berlin: Springer, 2003.

19

P. Simons.
Parts, A Study in Ontology.
Clarendon Press, Oxford, 1987.

20

M. Uschold.
Knowledge level modelling: Concepts and terminology.
Knowledge Engineering Review, 13(1), 1998.

21

A. Varzi.
Parts, wholes, and part-whole relations: The prospects of mereotopology.
Data and Knowledge Engineering, 20(3):259-86, 1996.