Ontology and Semantic Interoperability

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

1 Introduction

One of the major problems facing systems for Computer Aided Design (CAD), Architecture Engineering and Construction (AEC) and Geographic Information Systems (GIS) applications today is the lack of interoperability among the various systems. When integrating 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 software integration. Applications may use different terminologies to describe the same domain. Even when applications use the same terminology, they often associate different semantics with the terms. This obstructs information exchange among applications. To circumvent this obstacle, we need some way of explicitly specifying the semantics for each terminology in an unambiguous fashion. Ontologies can provide such specification. It will be the task of this paper to explain what ontologies are and how they can be used to facilitate interoperability between software systems used in computer aided design, architecture engineering and construction, and geographic information processing.

2 Languages and communication processes

Communication is an exchange of information about entities and relations between between a sender and a receiver. Information is formulated in some language. A language consists of symbols arranged in a well defined manner. The symbols of a language are not meaningful per se. The meaning of a symbol needs to be made explicit by specifying its intended interpretation, i.e., by specifying to which entity (entities) or relation it refers to.

We can think of information exchange as a sequence of distinct processes: (i) translating the symbols of the language in terms of which the sender expresses his information into a language that can be sent through a channel; (ii) sending the information encoded in this intermediate language through a channel to the receiver; (iii) translating the received symbols into symbols of a language in terms of which the receiver represents its information, and (iv) interpreting the symbols by identifying the entities and relations they refer to in the way intended by the sender. The (partial or complete) failure of any of these processes may result in a loss of information (Shannon and Weaver, 1949).

Spatial information, i.e., information about spatial entities and spatial relations between them, can be communicated, e.g., via intermediate languages such as natural language, graphical languages, and formalised computer languages. Today natural language is used mainly in communication between or to human beings. Natural language is used, for example, to communicate route directions, i.e., information about how to find a route in a spatial environment. Car navigation systems, for example, give route directions in natural language. Graphical languages are used in sketches and maps. Car navigation systems may give route directions not only in verbal form, but also use maps or graphical direction symbols on a screen. For the communication of spatial information between computers, languages of underlying data exchange formats such as *shapefiles*, or *dxf* are used. Particularly desirable in this context are languages which are standardised and whose specifications are available to the public, e.g., *GML*, or *VRML*.

Every language is characterised by its syntax and its semantics. The syntax concerns the symbols a language recognises and the rules which govern how to construct well formed sentences using those symbols. For languages used to communicate information, agreement about the rules of syntax is assumed as part of the accepted procedures between the communicating partners (Austin, 1975). In the specific case of spatial information, this agreement might mean that the sender uses grammatically correct natural language in verbal route directions, maps which conform to cartographic accepted procedures, or a VRML file with proper XML syntax. Deviations from a mutually accepted syntax complicates the decoding of the message (understanding) by the receiver and can lead to communication failure. For example, an error-tolerant web-browser might be able to repair some breaches of XML syntax, but will fail to read the transmitted information if other breaches occur.

The semantics of a language fixes the meaning of its expressions (symbols, terms, or sentences). Usually this is done by specifying interpretations for the language expressions in a given domain. The interpretation of a name is the individual it refers to. For example in most contexts in the English language the name 'The Eiffel Tower' refers to a specific steel construction in the centre of Paris. The interpretation of a predicate is a set of entities, e.g., the interpretation of the predicate 'is-blue' is the set of all blue things in the domain of interpretation. The interpretation of a n-ary relation symbol is a set of n-tuples of entities. For example, the interpretation of the relation symbol 'is-part-of' is the set of all ordered pairs (x, y) such that the individual x is a part of the individual y. If we constrain our attention, for example, to Tom's body parts, then the interpretation of 'is-part-of' contains ordered pairs like *(Tom's left thumb, Tom's left hand)*, *(Tom's left hand, Tom's left arm)*, *(Tom's left arm, Tom's* *body)*, etc.

The meaning of an atomic sentence determines its truth value: "Tom's arm is part of Tom's body" is true since Tom's arm is part of Tom's body, i.e., there is an ordered pair *(Tom's left arm, Tom's body)* in the relation denoted by the relation symbol 'is-part-of'. Atomic sentences can be combined to complex sentences using logical connections such as 'and' and 'or'. Let A and B be atomic sentences. The complex sentence 'A and B' is true if and only if A is true and B is true. Similarly, the complex sentence 'A or B' is true if and only if A is true or B is true.

3 Semantic heterogeneity

Communication obstructions arise from the fact that sender and receiver employ different languages for representing information internally. In the case of information systems these languages may have been established in different contexts and for a wide variety of purposes. As a result it may happen that the same symbol may have different meanings in different languages, or distinct symbols in different languages may have the same or overlapping meanings (Bishr, 1998; Vckovski et al., 1999). This semantic heterogeneity causes serious problems since it is often not clear how to interpret expressions properly in a communication process.

As a very simple example of semantic heterogeneity, consider the term 'tank'. In an information system used in a military context, it usually refers to a certain kind of armored vehicle. In an information system used to store information about zoological equipment, the term 'tank' refers to a kind of container which can hold water and serve as a habitat for fish. Now suppose that both an information system about armored vehicles and an information system about zoological equipment are used on a military basis and that the two information systems are to interoperate within a base-wide facility management system. In this case, it is not obvious how to interpret the expression 'three tanks'.

For a more complex example, consider the following. A typical problem within the planning process in Germany is the integration of data classified according to the ATKIS-OK-250 terminology system (provided by the German government) with data classified according to the CORINE land cover terminology system (provided by the European Community). To integrate these different data sets, we need to establish semantic relationships between the terms in the ATKIS and in the CORINE system. Comparing the two terminology systems reveals, for example, that ATKIS has a term *city-forest*, but CORINE has no term of the same meaning. A close match in CORINE is the term *sport-and-leisure-facilities* whose meaning overlaps but is not identical to that of ATKIS's *city-forest* (Visser et al., 2001). To determine whether a data item classified as *sport-and-leisure-facilities* according to the CORINE terminology can also be classified as a *city-forest* according to the ATKIS terminology, we need definitions that state the meaning of each term in some language that is more expressive than either ATKIS or CORINE.

To use terminology systems within a single domain or across domains in an unambiguous manner, it is important to make the semantics (i.e., the meaning) of the terms constituting the systems explicit. Assigning an explicit semantics to every terminology system enables us to interpret data items like '3 tanks' differently depending on whether the data is structured by a military terminology system or by a zoological terminology system. Similarly, explicit semantics for the CORINE and ATKIS terminologies are essential for integrating data entries like Auenwald-Leipzig is-a *sport-and-leisure-facility* (in CORINE) with data entries like Auenwald-Leipzig is-a *city-forest* (in ATKIS).

4 Ontologies

Ontologies are tools for specifying the semantics of terminology systems in a well defined and unambiguous manner (Gruber, 1993; Guarino, 1998). Ontologies are used to improve communication either between humans or computers by specifying the semantics of the symbolic apparatus used in the communication process. More specifically, Jasper and Uschold identify three major uses of ontologies (Jasper and Uschold, 1999): (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) and distinguish two major kinds of ontologies: logic-based and non-logic-based ontologies.

4.1 Logic based ontologies

A *logic-based* ontology is a logical theory (Copi, 1979). The terms of the terminology, whose semantics is to be specified, appear as names, predicate and relation symbols of the formal language. Logical axioms and definitions are then added to express relationships between the entities, classes, and relations denoted by those symbols. Through the axioms and definitions the semantics of the terminology is specified by admitting or rejecting certain interpretations.

Consider again the symbol 'is-part-of' interpreted as the (proper-) part-of relation as described above. An ontology can explicate the meaning of this symbol by stating that: (A1) if x is-part-of y the y is not a part of x, i.e., stipulating that the is-part-of relation is asymmetric, and (A2) if x is-part-of y and y is-part-of z then x is-part-of z, i.e., stipulating that the is-part-of relation is transitive. The statements (A1) and (A2) can be used as axioms of a logical theory of parthood. (A1) and (A2) specify meaning by excluding non-intended interpretations of the relation symbol 'is-part-of'.

Consider the relation *as-tall-as* which is constituted by ordered pairs like *(Tom, Jerry), (Jerry, Tom)*, etc., where Tom and Jerry are two people who are equally tall. Since axiom (A1) stipulates that the symbol 'is-part-of' must be interpreted as a relation that is asymmetric, it cannot be interpreted as the relation *as-tall-as*. This is because *as-tall-as* has the pairs *(Tom, Jerry)* and *(Jerry, Tom)* as members which taken together violate the asymmetry axiom (A1). The axioms of a logic-based ontology specify meaning by rejecting interpretations that do not conform with the intended use of the terms of the underlying terminology. Notice, that the technique of specifying the semantics of a terminology by *constraining* possible interpretations using an axiomatic theory is very general and not limited to a particular domain. For an extended discussion see (Guarino, 1998).

4.2 Non-logic-based ontologies

Often the semantics of terminology systems are specified using *non-logical* ontologies. Examples are ontologies stated in natural language as in the various ISO standards or in semi-formal languages such as UML.

Non-logical ontologies do not specify the semantics of a terminology system by constraining the permissible interpretations of the terms by means of logical axioms. An important class of non-logic-based ontologies are *standards*. A standard specifies the meaning of a terminology by fixing the interpretation of the terms with respect to a single, well defined, and fixed domain of interpretation. Disambiguity of terms is achieved since cases in which the same symbol has different meanings cannot occur and cases in which distinct symbols have the same meaning are avoided by agreeing on the use of terms.

Consider the standard specifying the semantics of the ATKIS terminology system. The semantics of the term 'forest', for example, is defined informally as a kind of vegetation area which has forest-plants or cultivated grass as vegetation, and in addition, has a size of at least ten hectares (this example was taken from (Visser et al., 2001)). This definition is very specific and meaningful only in the relatively narrow scope of the standard and with respect to the other terms specified within the standard. In a similarly specific way another standard specifies the intended meaning of the CORINE terminology.

Standards often appear where legislating bodies had the power to establish a common terminology for the scope of application of a law. Prototypical examples are ATKIS and CORINE. ATKIS is an established standard in the Federal Republic of Germany, and for official geographic data of the scale 1:25,000. Similar standards exist in nearly every country. With the CORINE project the European Commission defined a common terminology for land cover classifications in the area of the European Union to collect, coordinate and ensure the consistency of information about the environment and the natural resources in the Community.

Similar catalogues of shared terminology are established in numerous application areas of CAD. It is the economic pressure to share data in larger projects which drive this development. One arbitrary example is the body of rules defined jointly by the district heating industries of Germany, Austria and Switzerland (see, e.g., `http://www.agfw.de`). These rules are adopted by CAD systems for their layers for the utility industry. Some problems with standardisation in these application areas is the rapid technological progress, and the lack of obligation to follow agreed rules.

4.3 Meta-standards vs. reference ontologies

Consider the ATKIS and the CORINE terminology systems. Since the domains of interpretation of the two terminologies overlap, complex cases of semantic heterogeneity as discussed above may occur. Due to their informal and specific character, the standards specifying the semantics of the terminologies are not powerful enough to resolve those heterogeneities. For the integration of the two terminologies a third, more expressive terminology is required. The semantics of this terminology may be specified by a logic-based ontology, which then is called a *reference ontology.* The semantics of the reference terminology may be specified by a standard, which then it is often called a *meta*-standard.

Suppose we have a meta-standard or a reference ontology covering the terminology used in environmental planning. We can then establish semantic relationships between the terms in specific terminologies like ATKIS and CORINE and the terms defined in the broader terminology of environmental planning. The relationships between terms in ATKIS and CORINE are established, by translating first from one specific terminology to the broader terminology and then from the broader terminology to the other specific terminology. This strategy has been used with a rudimentary reference ontology in (Stuckenschmidt et al., 1999).

One advantage of the strategy of using a meta-standard or a reference ontology is that we do not need to establish direct links between all of the various terminology systems but only between each terminology system and the terminology specified by the relevant meta-standard or reference ontology. Also the terminology of the meta-standard or the reference ontology will ideally be formulated in expressive languages which enable us to make distinctions (e.g. between CORINE's sports-and-leisure-facility and ATKIS' city-forest) which cannot be made within the terminology systems.

(Meta)standard-based ontologies are useful in restricted domains and relatively homogeneous environments while the use of logic-based reference ontologies is more suitable for the integration of large terminologies in non-restricted domains and heterogeneous environments (Ciocoiu et al., 2000). Reference ontologies can be used to specify the semantics of rather general terminology systems and to integrate a broader variety of standards for at least two reasons: Firstly, the underlying semantics of logic-based ontologies is not limited to a single domain but is specified in a rather general manner by means of logical axioms. Secondly, due to the underlying logic the consistency of the ontology can be verified and intended and non-intended consequences be discovered. The second point is an important especially for large terminologies (Rector, 2003) and will be discussed in more detail in Section 4.4.

4.4 Logic-based reasoning

The reasoning facilities of the logical apparatus underlying a logic-based ontology can be used to compute consequences of the assumptions that have been made. For example, from the facts 'Tom's left thumb is part of Tom's left hand'

and 'Tom's left hand is part of Tom's left arm' a computer can, using axiom (A2), derive that Tom's left thumb is also part of Tom's left arm.

The reasoning facilities can also be used to discover non-intended consequences and inconsistencies. For example, in our ontology we might have: 'door handle part-of door' and 'door part-of house'. By (A2) we then have 'door handle part-of house'. This consequence might not necessarily be intended, since a door handle is not in the same sense a part of a house as the door, the roof, the walls, or the windows, which are parts which have a direct *function* for the house as a whole (Winston et al., 1987). If this consequence is unacceptable, then more complex notions of parthood, such as functional parthood or constitutional parthood, are required in our ontology (Artale et al., 1996).

The specification of the semantics of a terminology system by means of a non-logic-based ontology may be sufficient for human communication, since (i) humans understand natural language, and (ii) reasoning based axioms like (A1) and (A2) is part of human common sense reasoning (Davis, 1990). Computers, however, do not have this kind of background knowledge and built-in reasoning facilities. For this reason, ontologies that are intended as support for communication among computer programs or between humans and computers need to be specified in a language of formal logic which supports deductive reasoning and can be implemented on a computer.

4.5 Interoperability

There are at least two different ontology-based types of solutions to the problem of enabling different software applications to communicate: In the first type of solutions all applications *share a common terminology* in the communication process. The semantics of this shared terminology is often specified by a (meta) standard and all applications which adhere to the (meta) standard communicate using the same terminology in an unambiguous fashion. If an application internally uses a terminology that is different from the terminology of the standard then transformation mappings need to be established. If the application terminology has a well defined semantics (for example given by a different, more narrow standard) then semantic heterogeneity can be resolved by the human specialists who write the software that perform the transformation.

In the second type of solution is more flexible. Here applications use different terminology systems whose semantics are specified using logic-based ontologies. A broader terminology, whose semantics is also specified by a logic-based ontology, is used as an interlingua or reference terminology. Relationships between the terminologies are indirect: each terminology can be mapped into, or from, the reference terminology. Since the semantics of the more specific terminologies are specified using logic-based ontologies the mappings from and to the reference terminology can often be computed automatically (Stuckenschmidt et al., 2004). To enable computer programs to automatically generate transformations between different terminology systems is the core of the dream of the Semantic Web (Berners-Lee et al., 2001; Egenhofer, 2002).

With the growth of the Semantic Web the specification of the semantics of terminology systems using *description logic*-based ontologies has become popular. A Description Logic is a specific form of formal logic that can be run efficiently on a computer (Baader et al., 2002). In ontologies specified using a description logic, axioms like (A1) and (A2) can be represented and automatic reasoning can be performed without human assistance by a computer program.

5 Standards and reference ontologies for Spatial Information Systems

In the following sections we discuss potential uses of standards and reference ontologies for interoperating software applications in CAD, AEC, and GI processing. Note that, in the remainder, we use phrases like 'CAD, AEC, and GI systems' or simply 'spatial information systems' to refer to software systems used in CAD, AEC, and GI processing.

5.1 Spatial data standards and their limitations

In principle, both ontology-based solutions based on standards as well as solutions based on logic-based reference ontologies can be exploited to provide the foundations for systems that facilitate interoperability between the distinct software systems used in CAD, AEC, and GI processing. However, standardisation will be most successful in cases where software systems share common ground that can be made explicit and represented as a standard. This standard then enables interoperability by ensuring that all applications share a common terminology with an unambiguous semantics in communication processes, as described above.

For spatial information systems, this means that there can be a large degree of standardisation of the *spatial component* that can be exploited for facilitating interoperability. This is because the spatial components of these systems are based on terminologies that underly the processing and communication of information about spatial location. Already today data standards are applied quite successfully in the processing of this kind of spatial information.

Some prominent de-facto standards for communicating spatial information are the file formats *shapefile* and *dxf* (owned by the companies ESRI and Autodesk respectively). The specification of each file format defines a language with a terminology for expressing spatial information and rules of grammar that determine how to form well formed expressions. However, the provided terminology is rather narrow and limited to expressing relatively simple information about the geometry of spatial entities. Moreover the specification of the semantics is rudimentary and informal. Nevertheless, both file formats are accepted as standards for the communication of spatial information and most other vendors have enabled their products to directly read and write files in these formats.

It is important to recognise that, strictly speaking, the spatial components of CAD, AEC, and GIS only provide a means for processing and communicating information about the *location* of spatio-temporal entities. However information about location is only one aspect of spatio-temporal information. Spatio-temporal information covers information of all aspects of the wide variety of entities ranging from table-top scale (auto parts, computers) to large scale (rivers, continents), from human artifacts (cars) to natural phenomena (wetlands), from crisp entities with well defined boundaries (land parcels) to entities subject to vagueness and boundary indeterminacy (wetlands, mountains). We hold that to specify the semantics of a terminology system that is general enough to support the communication of information about entities characterised by a corresponding vast variety of different qualities and relations by means of a standard is very difficult, if not impossible.

Notice that this does not mean that there cannot be standards for attribute data. Standardised product catalogues are quite common, and ATKIS and CORINE certainly are standardised terminology systems for attribute data. Our point is that there is not likely to be a (meta-)standard that incorporates all (or sufficiently many) product catalogues used to annotate CAD and AEC data, or a meta-standard that incorporates all the (standardised) terminology systems used in AEC and standardised GIS terminologies including ATKIS and CORINE, etc. This is because the strength of a standard is that it is based on a well constrained terminology and the specification of the meaning of those terms within a limited and well defined domain. In such a framework there are no resources to deal, for example, with phenomena like vagueness, indeterminacy, and granularity in a way which is valid across different scales or different kinds of spatial entities.

To specify the semantics of a terminology system that is general enough to integrate a wide variety of different standards and to support the communication of information between heterogeneous sources such as CAD, AEC, and GI systems, a reference ontology is required.

5.2 Standards for the spatial component

Standardisation is sufficient for providing the basis for semantic interoperability among the *spatial* components of CAD, AEC, and GI systems. This is because the domain of interpretation of the terminology systems used to describe the spatial aspect of the entities represented in CAD, AEC, and GI systems is well understood, i.e., good mathematical models exist. As pointed out in (Chomicki and Revesz, 1999a; Kanellakis et al., 1990), the mathematical models that provide the semantics for any computer-implemented geometry language, no matter what dimension, are *semi-algebraic sets*: point sets forming lines, surfaces, volumes which are described using polynomial formulae in which only numbers that can be processed on a computer occur (Kanellakis et al., 1990). Thus the aim of any spatial data standard is to find a commonly accepted way of describing a well defined class of objects: semi-algebraic sets.

Notice however that, because of the particularities of computer arithmetic, the representation of semi-algebraic sets on a computer is far from trivial. However these problems are well known (Herring, 1991) and a variety of solutions have been proposed (Gueting and Schneider, 1995; Chomicki and Revesz, 1999b; Miller and Wentz, 2003). Eventually these solutions will find their way into a standard.

Standards are established in CAD and AEC as well as in GIS. An example for the previous is *Extensible 3D* (ISO, 2004), an XML-enabled format for the exchange of three-dimensional CAD data developed by the Web3D Consortium. An example for the latter is the *Simple Feature Specification* of the Open Geospatial Consortium (OGC) (Beddoe et al., 1999), which is also the basis for GML, the XML-enabled exchange format for two-dimensional geographic data. CAD/AEC and GIS standards differ not only in dimensions, but also in their primitives. CAD/AEC, and Extensible 3D in particular, offers boundary representations and parametric geometry (constructive solid geometry, CSG). GIS allow only boundary representations. The following paragraphs discuss some properties of standards taking OGC's simple feature specification as an example.

Simple features. The Simple Feature Specification introduces a terminology and specifies its semantics. Parts of the terminology are shown in Fig. 1. The terminology includes terms like 'geometry', 'point', 'line', etc. The standard organises these terms into a subsumption (is-a) hierarchy, i.e., the term 'geometry' subsumes the more specific terms 'point', 'curve', 'surface', etc. The interpretation of the term 'point' is specified informally as: "A zero-dimensional geometry and represents a single location in coordinate space. A point has a x-coordinate value and a y-coordinate value." (Beddoe et al., 1999, p. 2-4).

The semantics of the term 'curve' is specified as "a one-dimensional geometric object usually stored as a sequence of points, with the subtype of the curve specifying the form of the interpolation between the points. (Beddoe et al., 1999, p. 2-5). Currently there is only one term subsuming 'curve': 'LineString', which is interpreted as a linear interpolation between the points. The specification then continues to distinguish open and closed, and simple and non-simple (self-intersecting) curves, etc. (For more details see (Beddoe et al., 1999).)

Topological relations. Besides providing a terminology for expressing information about semi-algebraic sets (simple features), the OGC also provides a terminology for expressing information about *topological relations* between those sets. For that purpose the OGC standard utilises the nine-intersection model (Egenhofer and Franzosa, 1991). Using this formalism the standard provides a semantics for terms like *disjoint*, *touches*, *crosses*, *within*, and *overlaps*.

Let a and b be semi-algebraic sets denoted by geometric features according to the OGC standard, e.g., two areas or a line and an area, etc. We can identify the boundary of a, the interior of a, the complement of a, and for b respectively. The semantics of terms referring to topological relations that can hold between a and b is specified by characterising the intersection of the sets classified as interior, boundary, and complement with respect to a and to b. Between a and b a total of nine intersections can be built: the interior of a intersected with

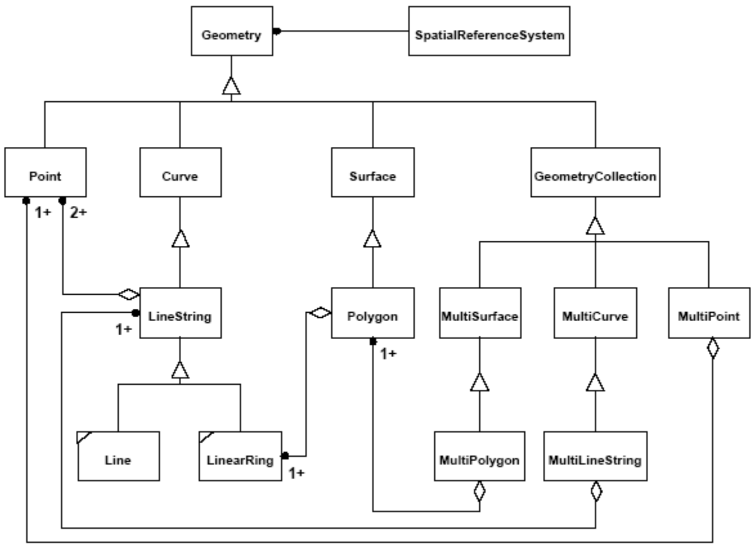

Figure 1: The geometry class hierarchy of OGC (from (Beddoe et al., 1999)).

the interior of b, the interior of a intersected with the boundary of b, and so on. The resulting intersections are sets, which may be empty or non-empty. Non-empty intersection sets may be of dimension 0 (i.e., points), 1 (lines), etc. The semantics of the term *disjoint*, is then, for example, defined as follows. If it holds that (i) an empty intersection of the interiors of a and b, (ii) an empty intersection of the boundary of a with the interior of b, (iii) an empty intersection of the interior of a with the boundary of b, (iv) an empty intersection between the boundary of a and the boundary of b, and (v) let the remaining five intersection sets be of any dimension, then, according to the standard, the relation that holds between a and b is the relation denoted by the term 'disjoint'.

Note that the nine-intersection model able to distinguish more relations than named in the standard. This causes semantic heterogeneity when different terminologies name the relations not covered by the standard differently (Riedemann, 2004).

Conformance testing. The question now is how to establish the relationship between terms in a terminology and abstract mathematical structures in a non-logic-based framework. OGC's answer to this problem is *conformance testing*. Since the standard is not based on logic, no logical axioms can be employed to specify the intended interpretation of a symbol like 'equal'. What the standard does provide are test procedures that partly enumerate the relation which is the intended interpretation of a relation symbol like 'equal'. These enumerations are called *test data*. Using test data it can be verified if a term like 'equal' used by a given application denotes the right relation. This is done by comparing the test data provided by the standard with the relation denoted by the term at hand.

To see how conformance testing works consider the symbol 'equal'. The relation denoted by this symbol is, according to the standard, supposed to contain ordered pairs of numbers like $(0, 0)$ and $(1.234, 1.234)$ but it should not contain pairs like $(2, 5)$ or $(0.00001, 0.0001)$. If the relation denoted by the application term 'equal' contains the pair $(0.00001, 0.0001)$ or fails to contain the pair $(1.234, 1.234)$ then this interpretation is not the one specified by the standard. Notice, however, that often relations denoted by terms like 'equal', 'greater-than', etc. are infinite or very large so that test data never can exhaustively ensure conformance with the standard.

Since the scope of the standard includes semi-algebraic sets, the relation denoted by the term 'equal' also holds between semi-algebraic sets. As in the case of numbers the standard provides test data for verifying the correct interpretation of the symbol 'equal' in the domain of semi-algebraic sets. The specification of the semantics of terms like 'disjoint', 'overlaps', etc. follows the same methodology.

OGC provides guidelines for conformance testing software implementing its simple feature specification (OGC, 1998). An implementation of an abstract specification (e.g., the relation denoted by a term like 'equal') is fed with a given test data set ("Joe's Blue Lake Vicinity Map") to verify its conformity with the specification of the standard. In this way a conformance test accomplishes alignment with the semantics of a standard.

Exchange formats. Together with the terminology and its semantics the OGC standard also specifies the grammar which describes how to form well formed expressions based on the given terminology. Using this language programs can read and write the well known binary and text formats to communicate, i.e., to export or import, information about geospatial features.

The OGC standard as meta-standard. The OGC standard is a meta-standard. Internally, each software application can describe semi-algebraic sets using quite different terminologies. Applications, for example, can use a language based on a polar coordinate system instead of a language based on Cartesian coordinates, or use a language of constraints on intersecting half-planes, etc. In the process of communication the internal terminology needs to be transformed into the terminology of the standard in a way that preserves the semantics. These translations are well known from mathematics (although not necessarily unique).

Notice, that the terminology used by the software internally may be richer than the terminology covered by the standard. A software could, for example, represent internally other types of curves than linearly interpolated ones. In such cases not all internal distinctions can be communicated from the sender to a receiver by means of the terminology provided by the standard.

OGC's geometry model of the simple feature specification is incorporated in the corresponding ISO norm (ISO, 2003), together with the topological operators. OGC's standard is one of several implementation specifications of OGC for making GIS interoperable. The standard is only two-dimensional and, hence, is insufficient for bridging the gap between CAD, AEC and GI systems. However, the principles of standardisation apply for all spatial information systems in the same way.

5.3 Limitations of today's data standards for CAD, AEC, and GIS integration

Besides the commonalities shared by CAD, AEC and GIS due to the spatial aspect of their data, there are also important differences. Several of them are discussed in detail elsewhere in this book. Differences concern which kind of spatial information is represented explicitly and and which kind of spatial information is omitted or represented only implicitly. We will mention here differences in dimensionality of the data and the capability to extract information about topology. In these areas, we need to develop data standards for CAD, AEC and GIS applications that go beyond standards that exist today.

Dimensionality. In the domains of CAD and AEC we typically process information about spatial entities of larger than geographic scale. Since information about location and extension in all three spatial dimensions is required, three-dimensional semi-algebraic sets are used to model spatial properties. The language used is the language of polynomials with three free variables for x, y, and z point coordinates.

GIS are designed to process information about entities of geographic scale. For this kind of entities it is often sufficient to process information of location and extension with respect to the surface of the Earth. For this reason two-dimensional semi-algebraic sets are used. The language to describe zero, one, and two-dimensional portions of the Euclidean plane is the language of (semi-algebraic) polynomials with two free variables for x and y point coordinates.

However, the surface of the Earth is not flat; it can be described in a complex mathematical language in three dimensions. Since the curvature of the Earth is relatively small, neglecting it is an acceptable simplification for areas of small geographic extent, e.g., in CAD and AEC. For areas of larger extent, curvature has to be considered. Hence, GIS represent the surface of the Earth in a cartographic projection onto a map plane, such that the third axis points (approximately) in the direction of the centre of gravity.

All cartographic projections show necessarily areal and angular distortions. The type of distortion at a specific location, as well as its size, depend on the actually chosen projection. Consequently, transformations between different projections are complex, but are necessary for integrating two data sets showing the same geographic area in different projections. Dealing with cartographic projections and transformations between them will be an essential part of standards that cover CAD, AEC and GIS applications.

Topology. Topology is implicit in any geometric representation. Hence, in a perfect mathematical world topology can be extracted from the information provided to specify the semi-algebraic sets. However, in computers we have to deal with finite representations of numbers, and finite precision of computations. Consider once again the relation denoted by symbol 'equal'. In the perfect world of mathematics a pair $(1.000001, 1)$ does not belong to the relation denoted by 'equal'. In a computer, however, where we can only distinguish a certain number of digits the numbers 1.000001 and 1 might be indistinguishable since 1.000001 has been truncated to 1.

If we extend that example to topological relations, we can easily show that often the intersection point of line a and line b computed using computer arithmetics is neither on line a nor on line b (Gueting and Schneider, 1995). This problem occurs if the coordinates of the mathematically correct intersection point of a and b cannot be represented in the language of the underlying computer arithmetic. In those cases a nearby point with representable coordinates is chosen as a result. This point, however, often does not lie on a nor on b.

Geometric representations that are the result of a construction process (e.g., a mechanical tool constructed using a CAD program) are different from geometries generated from measurement and observation data. Constructed geometries fit together nicely. Independently observed and measured geometries, however, are subject to measurement and observation errors. Different methods of observation and measurement yield data of different accuracy. Consequently the geometric representations of the same entity derived from data gained by different observation and measurement devices will be different semi-algebraic sets. Consequently, we cannot identify the different representations of the same entity using the predicate 'equal' which identifies sets only if they have the same members.

Similar problems occur for other topological relations such as disjoint, touches, and overlap. Given geometric representations of two entities generated by different observation methods, it is often the case that according to the representations built from one data set the entities are disjoint and according to representation built from another data set the entities overlap, while in fact both objects touch.

Both, errors caused by the specific character of computer arithmetics as well as measurement and observation errors particularly affect boundary representations used in early GI systems. For this reason software vendors felt a need for topological data models. In those data models topological relations are represented explicitly and not derived from the underlying geometric representation. GIS vendors started to include explicit representations of topology in their data models in the early nineties, with the first instance of TIGRIS, a system completely designed on basis of a topological data model (Herring, 1987), based on algebraic topology. Nowadays, all major GIS products include topology in their data model, and implement the standard set of topological relations (Beddoe et al., 1999; ISO, 2003).

Topology in three-dimensional space is more complex and problematic (Baer et al., 1979; Hoffmann, 1989). However, recent developments for three-dimensional

GIS based on algebraic topology (Breunig, 1996) might form the common mathematical foundation for developing a standard converging CAD and AEC with GIS.

6 Reference, domain, and top-level ontologies

After having discussed the use of standards in facilitating interoperability of CAD, AEC, and GIS software, we now focus on how to use logic-based ontologies for this purpose.

6.1 Domain ontologies as reference ontologies

Domain ontologies are ontologies that provide the semantics for the terminology covering a discipline. Since such terminologies are often large and complex they are potential fields of application for logic-based ontologies. Domain ontologies are prototypical candidates for serving as reference ontologies which facilitate the interoperability of software applications used within their domain.

Disciplines in which logic-based domain ontologies are quite common include Artificial Intelligence, medicine, biomedicine, and microbiology. Examples of medical domain ontologies are GALEN (Rector and Rogers, 2002a,b), SNOMED(CT) (Spackman et al., 1997), and the UMLS (Bodenreider, 2004). An example of a domain ontology for biomedicine and microbiology is the description logic based version of the GeneOntology (The Gene Ontology Consortium, 2001). In Artificial Intelligence the CYC-ontology is quite popular (Lenat and Guha, 1990).

Unfortunately there are only preliminary attempts to provide logic-based domain ontologies within the geo-domains (i.e., in domains in which CAD, AEC, and GIS are used for information processing). Examples are in (Grenon and Smith, 2004; Mark et al., 1999) for general ontologies of geographic categories, in (Sorokine and Bittner, 2005; Sorokine et al., 2004) for domain ontologies for ecosystems, and in (Feng et al., 2004) for a domain ontology for hydrology.

Logic-based geo-domain ontologies could provide semantic foundations for terminology systems used in the various geo-disciplines, for example for terms used to classify geo-political entities, or ecosystems, or to describe water-flow. A logic-based domain ontology for environmental planing, for example, may be used as reference ontologies for integrating the terms of specialised terminology systems, such as CORINE or ATKIS as described above. A logic-based domain ontology for architectural design and engineering could serve as reference ontology for specific terminologies underlying the usage of CAD systems and GI systems in this domain.

Building a domain ontology is an expensive and complex process (Rector, 2003). Recent research has shown that robust domain ontologies must be (Guarino, 1998; Gangemi et al., 2002):

1. developed rigorously using formal logic;

2. based on a well designed *top-level ontology.*

Above we have focussed on (1), we now consider (2).

6.2 Top-level ontologies

In contrast to domain ontologies, top-level ontologies specify the semantics for very general terms (called here top-level terms) which play important foundational roles in nearly every discipline. Top-level terms include relations like equal, is-part-of, connected-to, dependent-on, caused-by, instance-of, subclass-of, etc. These relations are used to structure information and define domain-specific terminology in geo-disciplines such as hydrology and environmental science, as well as in medicine, biology, and politics. For example, Germany is part-of the European Union, Canada is connected-to the United States, and South America is an instance-of continent. Within, e.g., an environmental planning domain ontology, we need to use top-level relations to regulate the usage of terms. For example, we might specify that: every instance of the class city-forest is a part of some instance of the class city.

Well designed domain ontologies use top-level ontologies as their foundation. This means that the semantics of the domain vocabulary is specified using top-level terms with an already well established semantics. One advantage of this approach is that top-level ontologies need to be developed only once and then can be used in many different domains. Another advantage is that a top-level ontology provides semantic links between the domain ontologies which are based on it.

The potential power of the methodology of building domain ontologies based on a well designed top-level ontology can be illustrated by considering the success of Egenhofer's formalization of the binary topological relations (a specific sub-collection of top-level notions) such as connected-to, overlaps-with, tangential-part-of, and so on (Egenhofer and Franzosa, 1991). Ten years after the introduction of Egenhofer's formalisation, the functionality based on this formalisation is part of all mainstream GIS and the terminology provided by Egenhofer is part of the OGC standard as discussed in Section 5.2. This could happen only because, despite the relatively abstract character of Egenhofer's formalisation, the relations treated in the formalism are familiar to researchers and practitioners in many domains. Egenhofer provided one component of a top-level ontology: a formal treatment of static topological relations. The Egenhofer formalism is the basis for uniform and semantically compatible strategies for representing and reasoning about topological data in environmental science, meteorology, urban planning, and other geo-disciplines.

6.3 Important components of top-level ontologies

Temporal aspects. Topological relations (and any other kind of properties and relations) are treated as time independent in today's CAD, AEC, and GI systems. This means that we can say that x and y are connected or that x is a part of y, but we cannot say *when* x and y stand in these relations. *Spatio-temporal* top-level ontologies will build on atemporal formalisms by constructing time-dependent spatial relations and properties. This is important for geo-domains as well as for AEC, because spatial properties and relations among entities in these domains change over time. The Czech Republic was not part of the European Union in 2001 but it is part of the European Union in 2004. The Auenwald in Leipzig was located in a singly connected region 100 years ago. Today it consists of multiple disconnected patches. Your car may have an old engine today and another newer tomorrow. Thus, often we need to say that x was a part-of y at time t_1 but x is no longer part of y at time t_2, or that x was located in y at t_1 but is no longer located in y at t_2.

Moreover, in disciplines such as hydrology, it is insufficient to collect and represent data only about enduring things (watersheds, rivers, etc.) and their changes over time (different size at different times, different water level at different times, etc.). It is critical also to collect and to represent data about the processes that cause those changes (e.g., soil erosion, water flow, etc.). A central component of a spatio-temporal top-level ontology will be a theory of the interaction between endurants (entities like watersheds that change over time) and perdurants (processes like soil erosion that unfold or develop over time).

Endurants and perdurants behave differently in time (Hawley, 2001; Gangemi et al., 2002; Masolo et al., 2004; Grenon and Smith, 2004; Bittner and Donnelly, 2004; Bittner et al., 2004a): Endurants are wholly present (i.e., all their current 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 current 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 ten years nevertheless you are the same person today you were ten years ago.

Perdurants, on the other hand, are extended in time in virtue of possessing different temporal parts which are characterised by different temporal extents. In contrast 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.

Individuals and classes. In geo-domain ontologies a logical theory of individuals and classes needs to provide the top-level notions that are needed for specifying the semantics of classification systems (Sorokine and Bittner, 2005; Sorokine et al., 2004). Particularly in geo-classifications at small scales, the distinction between classes and individuals (I am an individual, human being is a class) is often ignored. This in turn leads to an inconsistent usage of relations like part-of, instance-of, and subclass-of (is-a).

For example, in the Southeast Alaska Ecological Subsection Hierarchy (Nowaki et al., 2001) we find the assertion: Boundary Ranges Icefield is a subclass of Icefield. An ontological analysis reveals, however, that Boundary Ranges Icefield is an individual and Icefield is a class. Since subclass-of is a relation between two classes, Boundary Ranges Icefield cannot be a subclass of Icefield. By contrast, instance-of is a relation between an individual and a class. Thus, we can say that Boundary Ranges Icefield is an instance of the class Icefield. An example of the proper use of the subclass-of relation is the statement: Icefield is a subclass of Active Glacial Terrains (the class of all active glacial terrains). (See (Sorokine and Bittner, 2005) and (Sorokine et al., 2004) for an extended discussion.)

Such errors in the proper use of the top-level relations part-of, subclass-of, and instance-of make it impossible to achieve a consistent specification of the semantics underlying a classification (Guarino and Welty, 2000; Zhang et al., 2004). The resulting classification systems will be (at least partially) incompatible with other classifications. This in turn prevents exchanging data and interoperability at the level of software applications using those classifications. A logical theory of individuals and classes makes the distinctions between these different notions explicit and helps the domain specialist to use those notions in the appropriate manner. For a theory of this kind see for example (Bittner et al., 2004b).

6.4 Top-level ontologies for CAD, AEC, and GIS integration

Logic-based geo-domain ontologies are critical for integrating software used in CAD, AEC, and GI processing. Top-level ontologies facilitate the development of well formed domain ontologies. The following top-level notions are particularly important for the development of domain ontologies for integrating software used in CAD, AEC, and GI processing.

The notions of process and change (perdurants and the endurants they change) are critical in domains in which GIS have been used traditionally, for example in hydrology and in environmental science (Feng et al., 2004). 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 can be divided further into two major categories (Smith, 2003): 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. The following kinds of independent endurants can be distinguished: 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 similarly AEC systems have focussed on modelling aggregates, while fiat subdivisions such as land parcels were modelled 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 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 software systems used in CAD, AEC, and GI processing is the first step toward interoperability between those software systems.

7 Summary

In this paper we discussed how ontologies can be used to overcome the historic incompatibilities between software systems used in the domains of Computer Aided Design, Architectural Engineering, and Geographic Information Processing, and to facilitate the semantic interoperability among those systems.

We started with a discussion of the role of terminology systems in communication processes and how ontologies are used to specify the semantics of the terms in those systems. We distinguished two major kinds of ontologies: logic-based and non-logic-based ontologies. We also 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 non-logic-based 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 logic-based reference ontology is used as an interlingua which provides a means of transformation between the terminologies used by the different software applications.

For software used in the domains of CAD, AEC, and GI processing, we argued that the standard-based strategy is sufficiently powerful to facilitate the interoperability of the software systems for processing purely spatial data. We also argued that to achieve interoperability at the level of processing attribute data the more powerful and more flexible strategy of using logic-based 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 of CAD, AEC and GI processing, top-level ontologies need to be a foundational component of the reference ontologies.

Top-level ontologies describe notions that are so general that they are common to reference ontologies in any domain. For this reason they are of particular importance for the design of reference ontologies that are used to facilitate interoperability between domains as heterogenous as CAD, AEC, and GI processing.

Spatio-temporal top-level ontologies are critical for information processing not only in all the geo-disciplines and in architectural design and engineering, but more generally, in all disciplines dealing with any type of spatio-temporal phenomena. They facilitate the exchange of data and interoperability across different domains (e.g., geography, medicine, epidemiology, CAD, AEC) since they ensure that foundational spatio-temporal terms are used in a unified and semantically compatible manner.

Acknowledgements

The first and second author acknowledge gratefully financial support from the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the EU Network of Excellence in Semantic Datamining, and the Volkswagen Foundation Project "Forms of Life". The third author acknowledges an internal grant of the University of Melbourne.

References

Alessandro Artale, Enrico Franconi, and Nicola Guarino. Open problems for part-whole relations. In *International Workshop on Description Logics*, Boston MA, 1996.

John Langshaw Austin. *How to do things with words.* Clarendon Press, Oxford, 1975.

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

A. Baer, C. Eastman, and M. Henrion. Geometric modelling: A survey. *Computer Aided Design*, 11(5):253–272, 1979.

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.

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

Y. Bishr. Overcoming the semantic and other barriers to gis interoperability. *International Journal of Geographical Information Science*, 12(4):299–314, 1998.

T. Bittner and M. Donnelly. The mereology of stages and persistent entities. In R. Lopez de Mantaras and L. Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence*, pages 283–287. IOS Press, 2004.

T. Bittner, M. Donnelly, and B. Smith. Endurants and perdurants in directly depicting ontologies. *AI Communications*, 14(4):247–258, 2004a.

T. Bittner, M. Donnelly, and B. Smith. Individuals, universals, collections: On the foundational relations of ontology. In A.C. Varzi and L. Vieu, editors, *Proceedings of the third International Conference on Formal Ontology in Information Systems, FOIS04*, volume 114 of *Frontiers in Artificial Intelligence and Applications*, pages 37–48. IOS Press, 2004b.

O. Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*, 2004.

Martin Breunig. *Integration of Spatial Information for Geo-Information Systems*, volume 61 of *Lecture Notes in Earth Sciences*. Springer, Berlin, 1996.

J. Chomicki and P.Z. Revesz. Constraint-based interoperability of spatiotemporal databases. *Geoinformatica*, 3(3):211–243, 1999a.

J. Chomicki and P.Z. Revesz. A geometric framework for specifying spatiotemporal objects. In *Proc. Sixth International Workshop on Temporal Representation and Reasoning*, 1999b.

M. Ciocoiu, M. Gruninger, and D.S. Nau. Ontologies for integrating engineering applications. *Journal of Computing and Information Science in Engineering*, 2000.

I.M. Copi. *Symbolic Logic*. Prentice Hall, 1979.

E. Davis. *Representations of Commonsense Knowledge*. Morgan Kaufmann Publishers, Inc., 1990.

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.

Max J. Egenhofer and Robert D. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2): 161–174, 1991.

C-C. Feng, T. Bittner, and D.M. Flewelling. Modeling surface hydrology concepts with endurance and perdurance. In M. J. Egenhofer, C. Freksa, and H. J. Miller, editors, *Proceedings of GI-Science 2004*, LNCS 3234, pages 67–80, 2004.

A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with DOLCE. In *Proceedings of EKAW*, Siguenza, Spain, 2002.

P. Grenon and B. Smith. SNAP and SPAN: Prolegomenon to geodynamic ontology. *Spatial Cognition and Computation*, 2004.

T. Gruber. A translation approach to portable ontology specification. *Knowledge Acquisition*, pages 199–220, 1993.

N. Guarino. Formal ontology and information systems. In Nicola Guarino, editor, *Formal Ontology and Information Systems, (FOIS'98)*. IOS Press, 1998.

N. Guarino and C. Welty. A formal ontology of properties. In R. Dieng and O. Corby, editors, *Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management*, LNCS. Spring-Verlag, 2000.

R.H. Gueting and M. Schneider. Realm-based spatial data types: The rose algebra. *VLDB Journal*, 4:100–143, 1995.

K. Hawley. *How things persist.* Oxford : Clarendon Press, 2001.

John R. Herring. Tigris: Topologically integrated geographic information system. In *Auto-Carto 8*, pages 282–291, Baltimore, 1987. ACSM - ASPRS.

John R. Herring. The mathematical modeling of spatial and non-spatial information in geographic information systems. In David M. Mark and Andrew U. Frank, editors, *Cognitive and Linguistic Aspects of Geographic Space*, volume 63 of *Nato ASI Series D*, pages 313–350. Kluwer, Dordrecht, 1991.

Christoph M. Hoffmann. *Geometric and Solid Modeling.* Series in Computer Graphics and Geometric Modeling. Morgan Kaufmann, San Mateo CA, 1989.

ISO. Geographic information - spatial schema. International Standard ISO 19107, ISO, 2003.

ISO. Information technology computer graphics and image processing extensible 3d (x3d). International standard 19775:200x, ISO/IEC, 2004.

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.

P.C. Kanellakis, G.M. Kuper, and P.Z. Revesz. Constraint query languages. In *9th ACM PODS*, pages 299–313, 1990.

Douglas B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project.* Addison-Wesley, Reading, Massachusetts, 1990. Lenat, D. B. Guha, R.

D. Mark, B. Smith, and B. Tversky. Ontology and geographic objects: An empirical study of cognitive categorization. In C. Freksa and D. Mark, editors, *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, number 1661 in Lecture Notes in Computer Science, pages 283–298. Springer Verlag, 1999.

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

Harvey J. Miller and Elizabeth A. Wentz. Representation and spatial analysis in geographic information systems. *Annals of the Association of American Geographers*, 93(3):574–594, 2003.

Gregory Nowaki, Michael Shephard, Patricia Krosse, William Pawuk, Gary Fisher, James Baichtal, David Brew, Everett Kissinger, and Terry Brock. Ecological subsections of southeastern Alaska and neighboring areas of Canada. Technical Report R10-TP-75, USDA Forest Service, Alaska Region, October 2001.

OGC. Conformance test guidelines for OpenGIS simple features specification for SQL, revision 1.0. OpenGIS Project Document 98-046r1, Open GIS Consortium, Inc., 1998.

A.L. Rector and J.E. Rogers. Ontological issues in using a description logic to represent medical concepts: Experience from GALEN: Part 1 - principles. *Methods of Information in Medicine*, 2002a.

A.L. Rector and J.E. Rogers. Ontological issues in using a description logic to represent medical concepts: Experience from GALEN: Part ii - the GALEN high level schemas. *Methods of Information in Medicine*, 2002b.

R. Rector. Modularization of domain ontologies implemented in description logics and related formalisms including OWL. In *Proceedings of the international conference on Knowledge capture*, pages 121–128, 2003.

Catharina Riedemann. Towards usable topological operators at GIS user interfaces. In F. Toppen and P. Prastacos, editors, *7th Conference on Geographic Information Science (AGILE '04)*, pages 669–674, Crete, Greece, 2004. Crete University Press.

Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Illinois, 1949.

B. Smith. Basic formal ontology. Technical report, Institute for Formal Ontology and Medical Information Science, University of Leipzig, 2003.

A. Sorokine and T. Bittner. Understanding taxonomies of ecosystems: a case study. In P. Fisher, editor, *Developments in Spatial Data Handling*, pages 559–572. Springer Verlag, Berlin, 2005.

Alexandre Sorokine, Thomas Bittner, and Chris Renschler. Ontological investigation of ecosystem hierarchies and formal theory for multiscale ecosystem classifications. In *Proceedings of GI-Science 2004*, 2004.

K.A. Spackman, K.E. Campbell, and R.A. Cote. SNOMED RT: A reference terminology for health care. In *Proceedings of the AMIA Annual Fall Symposium*, pages 640–4, 1997.

H. Stuckenschmidt, U. Visser, G. Schuster, and T. Voegele. Ontologies for geographic information integration. In *Proceedings of the workshop intelligent methods fro handling enviromental information: Special aspects of processing space and time*, 1999.

Heiner Stuckenschmidt, Frank van Harmelen, Paolo Bouquet, Fausto Giunchiglia, and Luciano Serafini. Using C-OWL for the alignment and merging of medical ontologies. In O. Hahn, editor, *Proceedings of KR-MED 2004: First International Workshop on Formal Biomedical Knowledge Representation*, 2004.

The Gene Ontology Consortium. Creating the gene ontology resource: Design and implementation. *Genome Res*, 11:1425–1433, 2001.

Andrej Vckovski, Kurt E. Brassel, and Hans-Jörg Schek, editors. *Interoperating Geographic Information Systems*, 1999.

U. Visser, H. Stuckenschmidt, H. Wache, and Voegele. Using environmental information efficiently: Sharing data and knowledge from heterogeneous sources. In C. Rautenstrauch and S. Patig, editors, *Environmental Information Systems in Industry and Public Administration*, pages 41–74. Hershey, PA, IDEA Group, 2001.

M.E. Winston, R. Chaffin, and D. Herrmann. A taxonomy of Part-Whole relations. *Cognitive Science*, 11:417–444, 1987.

S. Zhang, P. Mork, and O. Bodenreider. Lessons learned from aligning two representations of anatomy. In O. Hahn, editor, *Proceedings of KR-MED 2004: First International Workshop on Formal Biomedical Knowledge Representation*, 2004.