DESIGNING A CARTOGRAPHIC ONTOLOGY FOR USE WITH EXPERT SYSTEMS

Richard A. Smith

University of Georgia, Geography Department, Athens, GA 30602 rasmith@uga.edu

KEY WORDS: Ontology, Cartography, Expert Systems, Map

ABSTRACT:

The democratization of mapping represents a positive trend for society, however within this movement, little focus to cartographic quality has been paid. The result is often poorly presented maps which fail to convey the information intended. Software packages offer very little, if any, guidance to the user when it comes to offering cartographic assistance such as symbolization recommendations. Additionally, users often do not read software manuals, and, instead, opt to solve the problem by trial and error or ask an available expert. The combination of scarce cartographic documentation and the resistance of users to use documentation present an opportunity for expert systems to provide better assistance through a non-passive help system. The paper presents the development of ontologies and their potential contribution to the construction and successful execution of an expert system. Traditionally, expert systems learned the situation through a dialog with the user after which it would infer an expert answer. A problem with this model is that if the users do not have sufficient knowledge of cartography to answer these questions, the expert system will be less effective at helping. If the expert system can understand the situation without querying the user, it can make proactive inferences, thus, reducing the burden on the user and reducing the amount of time spent interacting with the expert system. To effectively replace user input to an expert system, it must be able to understand the concept of a map. The construction of an ontology will provide this understanding.

1. INTRODUCTION

The primary function of a map is to communicate; therefore, it is important to clearly convey information graphically in order to maximize the likelihood of the map successfully communicating to an audience. Designing such a cartographically well-designed map is a time consuming process consisting of selection and preparation of data, choosing visual representations (such as color, line weight, symbols, patterns), labeling and layout. The map creator must possess sufficient cartographic knowledge to design a map that fulfills its primary function. Many maps produced today often do not effectively convey the intended information due to the map maker's lack of cartographic training. The wide-spread use of personal computers and prevalence of map-making software allows anyone to create a map regardless of experience. Software does not provide feedback or suggestions on cartographic choices, thus, users may not even realize that their map product is sub-par, cartographically speaking. If expert cartographic knowledge can be collected and distributed in an expert system, map makers of all proficiency levels can receive guidance from the expert system improving their ability to produce a well-designed map. Before the expert system can be constructed, however, a common knowledge framework needs to be constructed to facilitate communication of cartographic knowledge between the knowledge engineer, cartographer, expert system and map makers. The basic premise of this research is that the creation of an ontology will provide the requisite common knowledge framework for cartographic knowledge acquisition for use by an expert system.

1.1 Background

Many maps produced today often do not effectively convey the intended information due to a lack of training on cartographic

knowledge. The wide-spread use of personal computers and prevalence of map-making software allows anyone to create a map with a few clicks of a button. While this democratization of mapping is seen as a positive trend for society (Mattmiller 2006), since the beginning of computerized cartography it has been noted that "cartographic monstrosities" (Monmonier 1984) have been caused by users with subpar (or no) cartographic training (Muller 1983).

The combination of scarce cartographic documentation and the resistance of users to use documentation present an opportunity for expert systems to provide better assistance through a non-passive help system. An expert system is a computer program that emulates the decision-making ability and knowledge of a human expert through the passing of facts to and receiving expertise from the expert system (Giarratano and Riley 2005).

A cartographic expert system is currently in development aimed at assisting users in designing a cartographically-sound general reference map. The development of an expert system was chosen over a typical linear algorithmic approach because the cartographic design process is an ill-structured problem which algorithms are not well suited to handle. The expert system contains rules about cartographic design and reasoning and explanation faculties designed to partially automate the cartographic design process and explain its reasoning to the user to teach cartographic techniques. In essence, the expert system will fill the role of the expert where an expert is unavailable.

It became clear during the development of the expert system that without a formalized way for computers to "understand" the composition of a map, then it would be unable to intelligently apply cartographic rules thus leaving the quality of maps impaired. Much of the mapping process is intuitively understood

by humans but cannot be understood by a computer without the process being explicitly defined and bounded. Additionally, we cannot simply start coding the mapping process in a computer without first defining the concept of a map. In the same way that we would first explain the concept of a map to a person unfamiliar with the concept, we must do the same with a computer. Constructing an ontology covering the concept of a map is the first step in satisfying this knowledge deficiency for the expert system and is the focus of this research.

1.2 Overview of Ontologies

An ontology is a formal specification of a shared understanding of a knowledge domain that facilitates accurate and effective communications of meaning (Gruber 1993; Agarway 2005). In other words, an ontology is composed of a taxonomic structure which contains objects and their relationships to other objects within a particular domain of knowledge. In use, the ontology is a framework on which concepts can be formalized.

In understanding what an ontology is and its function, four key ideas need to be understood. The first idea, *concept*, means an abstraction of the world that is to be modeled. The world is abstracted as objects which have properties and relationships to other objects inside and outside the domain of knowledge being abstracted. Additionally, if we choose to specify constraints, we can construct an ontology that validates objects and their relationships.

The second idea, *formal*, means that the ontology is specified in a structured, logical, unambiguous fashion allowing for machine processing. This is accomplished through the use of a structured language built to be handled consistently and maintain structure.

The third idea, *explicit*, means that the ontology's concepts are clearly defined in a carefully selected vocabulary. The vocabulary should utilize words that describe the concept in the most precise manner and are commonly used by domain experts. Using unfamiliar or controversial words as the basis for the concept is counter-productive to the purpose of ontology construction.

The fourth idea, *shared*, means that the ontology represents concepts agreed upon by a community of experts. The word *shared* also refers to the formal language utilized as the ontology's dissemination vehicle. Using a formal language with a common syntax allows for the ontology to be shared between systems and users without losing structure or meaning.

Four broad categories of ontologies have identified by Guarino (1997): 1) top-level ontologies, 2) domain ontologies, 3) task ontologies, and 4) application ontologies. A top-level ontology covers a wide range of concepts independent of an individual problem. Top-level ontologies are concerned with abstract concepts and often serve as a foundation for more specific ontologies and sharing of knowledge across domains. Domain ontologies represent objects and their relationships within a specific domain of knowledge. The domain ontology is tasked with defining the vocabulary and relationships of the domain. Task ontologies deal with a specific task or activity within a certain domain. It is unlikely that the knowledge represented in the task ontology will be useful outside of its specific domain.

Application ontologies link the domain and task ontologies acting as a mediator.

The map ontology will be constructed as a domain ontology. The domain ontology was chosen because a top-level ontology would have to be too general for our purposes, and the concept of a map is already (broadly) defined in some top-level ontologies. Our ontology will tie into the top-level ontology's map concept but provide more specific details.

In the context of this research, the ontology will serve as a specification of the concept of a map to facilitate knowledge acquisition from experts and design of the expert system to enable it to understand and intelligently apply the acquired expert knowledge. Development of an ontology has benefits to expert systems development. The first benefit, re-usability, shares it name with a major tenant of effective computer programming; re-use as much as possible and do not "re-invent the wheel". An exhaustive search has been performed by the author in the attempt to find an existing cartographic ontology, but the search yielded no immediately useable domain ontologies. The lack of immediately useable cartographic domain ontologies prompted this research, thus, this research will create the needed cartographic ontology and will be shared for future re-use and expansion by others.

The second benefit of ontology development, knowledge acquisition, speeds the acquisition of knowledge as the ontology provides a basis for knowledge engineers to ask questions of the domain expert.

Reliability is the third benefit and provides the automation of consistency checking, which would be difficult if done manually as ontologies can become quite complex. If assertions in the ontology are inconsistent (for instance, stating that a dog is both a type of plant and animal), then a computer reasoning with this inconsistent ontology may produce erroneous conclusions.

This research aims to have the ontology enable machine understanding of cartographic concepts with respect to the construction of maps. The development of this ontology will allow computers and cartographers to share their knowledge, thus making this knowledge accessible to a larger audience.

2. METHODOLOGY

Constructing an ontology is a multi-step process that is necessarily iterative and difficult to progress through linearly. In an effort to streamline construction of the ontology, multiple ontology development methodologies were researched and evaluated. Many of the ontology building methodologies shared similarities, and in the end, METHONTOLOGY (Fernandez, Gomez-Perez, and Juristo 1997) was the chosen methodology as it incorporated the best practices elicited by many of the other methodologies.

In METHONTOLOGY, the development life-cycle is composed of specification, conceptualization, formalization, integration, implementation, evaluation and documentation. The ontology life-cycle does not specify the order in which the tasks should be undertaken. During the initial iteration of the ontology development, some tasks will naturally come before others, such

as specification of the problem before implementation, however as the ontology is iteratively constructed, task order becomes less relevant. Each of the tasks will now be discussed as they were completed for creation of this map ontology followed by a discussion of the resultant ontology in the results section of this paper.

2.1 Specification

The process of creating an ontology initially involves consulting references (such as texts and experts) to a) locate previous attempts at constructing a related ontology, and b) frame the domain to be conceptualized. The specification task requires a document to be drafted in natural language that includes such information as the purpose, level of formality, scope, vocabulary, competency questions and sources of knowledge. This document guides the development process and allows for brainstorming and then bracketing of the domain to be modeled. This specification document provides guidance throughout the ontology development process by keeping the problem bounded and on topic. Multiple text sources and experts were consulted during the specification phase to define and bound the problem. Each section of the specification document will now be discussed.

Purpose: The purpose of the ontology is to describe the constructs required to build a map. It will assist a knowledge engineer in eliciting cartographic expertise from cartographers and translating this information into an expert system with the help of a computer scientist. The expert system will use the ontology to determine appropriate selection, placement and symbolization of graphic variables. The ontology must also describe the interaction between the map, the objects which compose it, and the real-world phenomenon that it will represent. Ultimately, the scenario in which the ontology will be used is to create the basis for questions to elicit knowledge from cartographers and to assist in building the framework for the expert system's facts and rules.

The audiences of this ontology are cartographers, computer scientists and an expert system. The ontology should be understandable by a cartographer with some study and help of a knowledge engineer. The ontology needs to be sufficiently formal for the computer scientist to encode the knowledge in a way in which the expert system can read. For the expert system, the ontology must include even the most basic knowledge as it cannot be assumed that the expert system has any pre-existing knowledge in this domain.

Level of Formality: Four types of ontologies have been distinguished by Uschold and Gruninger (1996) and are: highly informal ontologies, semi-formal ontologies, formal ontologies and rigorously formal ontologies. A highly informal ontology is expressed loosely in a natural language, such as a narrative written in the English language. This type of ontology is a poor fit as computers are not able to reliably and easily interpret such an ill-structured document and ambiguous language. A semi-informal ontology is expressed in a more structured form of a natural language with a well defined vocabulary, however, this type of ontology, too, is a poor choice much for the same reasons the highly informal ontology type is a poor choice. A semi-formal ontology is written in an artificial and formally defined

language that is machine interpretable because of the strict structure and language requirements. A rigorously formal ontology is a precise document formed by well defined terms, semantics, theorems and proofs.

For this research, the formal ontology was the chosen type of ontology. Because both a human and computer need to be able to easily interpret the ontology, the formal ontology type was the best fit of the four.

Scope: Scope defines the context and domain in which the ontology will operate. It is important to have a bounded scope and domain as to keep the ontology focused and lean. The map ontology is bounded to a tangible or virtual map that can be viewed by many people. Additionally, the ontology will only cover the basic structure of a map and the representation of space thereon. The level of granularity is purposely coarse as to allow flexibility in subsequent contributions.

Vocabulary: To assist in determining the scope and requirements of the ontology, a list of terms of objects or concepts a cartographer might use when designing a map was compiled. Once the terms were collected, they were classified into logical groups and served as an initial structure of the ontology. These terms were also used as a basis for determining how the elements interact with each other to compose a map, and formed a basis for questions to start building the expert system's knowledgebase.

Competency Questions: The role of the competency questions is to define the scope of the ontology and serve as a set of guiding thoughts when constructing the ontology. The ontology is deemed successful if it can be used to satisfactorily address the questions. For the map ontology, four competency questions were established:

- What are the components of a map?
- How do map components relate to compose a map?
- How do objects in the real world become transformed to representations on the map?
- What visual elements are applied to representations of objects?

Sources of Knowledge: Multiple cartographers, maps, cartography texts, and ontologies were consulted in determining the terms and scope of the ontology.

2.2 Knowledge Acquisition

The knowledge acquisition step extends across the entire ontology development life cycle. Initially, knowledge acquisition is used to create the initial specification of the ontology. This can be done through brain storming, interviews, literature review or other useful methods. Later in the ontology development life cycle, knowledge acquisition provides the ontology validation service. In this research, the knowledge acquisition step progressed in parallel with the initial specification step and ran concurrently throughout the remainder of the ontology development task.

Four relevant sources of information and knowledge were identified for development of the map ontology: expert cartographers, maps, cartography texts and ontologies. Each of these sources of information were tapped throughout the

construction of the ontology and were used to validate the ontology. Additionally, each source was used to corroborate knowledge acquired from the other sources.

A search for existing map ontologies bore few results and only one that provided enough detail to be useful as a source of information. The found ontology provided a skeleton for a basic domain cartographic ontology (Iosifescu-Enescu and Hurni 2007) and yielded some re-useable portions. Even though the ontology was not complete, it offered an initial vocabulary and structure for development of our map ontology.

Interviews were conducted with cartographers to build upon the preliminary specification document and vocabulary. The concept and construction of a map were discussed to determine what knowledge and information a cartographer must possess to construct a map. The specification document and vocabulary were then expanded upon using informal analysis of texts to flush out the structure and to corroborate the knowledge gained from the interviews. The knowledge gained up to this point was then evaluated against a plethora of maps as a sanity check and to identify areas that had been ill-defined. Once satisfied that the knowledge was valid, an initial ontology was conceptualized (see next section). The final step was taking the vocabulary, specification document and constructed ontology back to the expert cartographers for another round of discussions and evaluation of the ontology.

2.3 Conceptualization

The conceptualization step deals with the structuring of the domain knowledge discovered in the specification step. The vocabulary was expanded to a complete glossary of terms which included the acquired vocabulary, verbs and properties. The glossary was compiled from the specification document and initial vocabulary and then was structured more formally for use in developing the ontology. The conceptualization step started and ran concurrently with the second half of the specification step.

2.4 Integration

The integration step proposes the use of other ontologies to speed the construction of the new ontology. The use of a metaontology, such as OpenCyc (CyCorp 2010) is recommended as they cover a large range of topics and provide a standard set of terms used by other ontologies. This step started concurrently with the initial specification step and yielded one domain ontology (as previously mentioned) that provided some basis for a starting point.

2.5 Implementation

The implementation step is where the ontology is actually developed using a well structured format. In our case, we chose to use the web ontology language (OWL) which is recommended by the W3C (W3C 2004) and requires structure and semantics on objects and supports the specification relationships between objects specified in the language. OWL and RDF are well established, widely used and can support a formal ontology. OWL is stored in XML format which can be consumed by the expert system.

To build the ontology, the software program Protégé (http://protege.stanford.edu/) a free, open-source ontology editor was used. Protégé fully supports OWL and provided the functionality required to create this ontology. Additional functionality of Protégé that was useful was lexical and syntactic checking, search functionality and visualization.

2.6 Evaluation

Evaluation step is where a technical judgment of the ontology is performed with respect to the domain of knowledge being modeled in the ontology. This step requires verification of the correctness and the validation of the structure of the ontology. The ontology was verified by expert cartographers and against maps. The structure was validated by the software program Protégé. The evaluation step was performed throughout the ontology development life cycle in an effort to catch incomplete, redundant and inconsistent knowledge and structures.

2.7 Documentation

The documentation step requires explanatory documents to be created throughout the ontology development life cycle. The purpose of documentation is to record decisions and the information those decisions were made with respect to for future reference and justification. The documentation for the map ontology takes many forms such as the specification document, vocabulary, internal explanatory documents, reference managers and documentation within the ontology itself.

3. RESULTS: THE MAP ONTOLOGY

In this section, the map ontology will be discussed first as a whole, and then by the eight top-level classes of objects. In OWL, objects represent things. Before we discuss the ontology, however, some terms and ideas need to quickly be discussed.

3.1 Formal Ontology Concepts

Class: A class represents tangible or intangible things that exist. Classes state the requirements for which a thing would be considered to be a type. For instance, if we had a class to represent a dog, the requirements would be (a) animal, (b) has four legs, (c) barks and (d) is warm blooded. A dog that exists would then be considered to be a type of dog as specified by the dog class.

Property: A property is a relationship between two classes. A property may be used to define restrictions between classes as well, such as restricting the cardinality.

Inheritance: In an ontology, classes are placed in a hierarchical structure with respect to whether a class is a subclass (child) of another superclass (parent) class. A child class must have all of the properties of its parent class and must be more specific of a class than its parent without exception.

3.1 Top Level Classes of the Map Ontology

The top-level (most general) classes in the map ontology are: Attribute, Graphic, Layout Element, Map, Map Projection,

Production Medium, Spatial Phenomenon and Visual Variable. These eight classes encompass all of the concepts and information required to make a map. It is important to keep in mind the purpose and scope of the ontology when evaluating the grouping of concepts in the ontology.

The Map class is the focus of this ontology and will be the focus of the discussion. A map (represented by the Map class) is a communication device that displays representations of mainly spatial phenomenon. The spatial phenomenon are described by attributes and represented by graphics, such as points, lines, polygons and rasters. The graphics are then visualized using such concepts as color and opacity and placed on a medium such as paper using a map projection. Attributes are visualized with labels and are also placed on the production medium. Figure 1 illustrates how these top-level classes and a few important child classes work together to construct the concept of a map. Note that only the top-level classes and necessary subclasses are shown in Figure 1. Many other classes exist in the ontology and are discussed below.

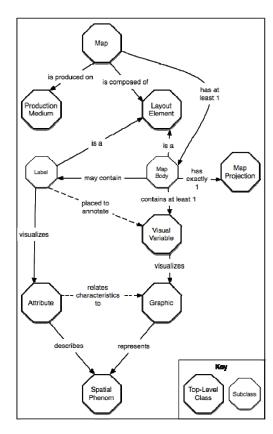


Figure 1. Taxonomy of the Map Ontology

Map: All other classes in the ontology will related directly or indirectly to the map class, which is the container for the concept of a map. It represents the most basic concept of a map. In order for something to be considered a map, it must communicate something about a spatial phenomenon. This communication can range from a simple map from a store to a house showing only 3 streets, two buildings and three labels to a very complex map showing the economic interactions between countries. Whether

the map is simple or complex, all maps have a basic set of requirements before they are considered to be a map. These requirements are that they show a representation of a spatial phenomenon (bona-fide or fiat) on some medium. Whether the map communicates effectively is irrelevant as even poorly designed maps were still intended to communicate a spatial phenomenon.

In the map ontology, the Map class has two children: Reference Map and Thematic Map. These two types of maps meet the basic set of requirements discussed above, but display different facets of spatial phenomenon. Reference maps display objects from the environment and show features of spatial phenomenon. Thematic maps present a graphic theme about a physical or abstract subject. Thematic maps are further subdivided as qualitative and quantitative depending on which type of attribute is being mapped.

Map Projection: The map projection class represents the task of placing the representations of spatial phenomenon on a surface for visualization. Any surface for visualization requires a projection whether it is a flat piece of paper or a virtual globe. Every map body (some maps containing more than one map body) must have exactly one map projection making the map projection a requirement for a map.

In the map ontology, properties of map projections were chosen as the delimiter for child classes. The five properties widely accepted are: azimuthal, conformal, equal area, equidistant and compromise. The compromise map projection contains elements from two or more of the other properties, but cannot hold all four true at once.

Production Medium: As the scope for the map ontology is restricted to physical or virtual maps that can be viewed by multiple people, the map must be produced on viewable medium. This class is a required part of the concept of a map. The two subclasses of Production Medium are Paper and Digital.

Layout Element: The layout element class represents the objects that construct a map. The layout element class has eleven subclasses: Map Body, Legend, Directional Indicator, Metadata, Neatline, Scale, Title, Lable, Ancillary Text, Ancillary Object and Graticule. Of the eleven classes, only one Map Body is required by the Map class. While it is preferable for additional layout elements to be used on map to provide supporting information, the scope of the ontology set forth earlier dictates that the most basic type of map be conceptualized.

Most of these subclasses are self explanatory based on their name, but we will take a moment to discuss ones that might be unfamiliar or unclear.

Ancillary Text refers to explanatory text displayed on the map. Such text could be descriptive or flavor text. Ancillary Object refers to objects such as graphs, photographs and videos placed on the map. Directional Indicator is more commonly known as a north arrow, however, not all maps specify direction using North, or any of the other cardinal directions. To maintain a consistent level of granularity across the ontology, a north arrow should be considered a subclass of Directional Indicator.

Spatial Phenomenon: The Spatial Phenomenon class represents all bona-fide and fiat objects and are the objects the map is made to communicate. It is important to note that the map does not contain spatial phenomenon, but abstract representations. These abstract representations are created by humans and are stored in graphic representations (discussed below). For the purposes of mapping, spatial phenomenon are divided into two child classes relating to the way in which they are abstracted: discrete and continuous.

Discrete spatial phenomenon have defined boundaries (precise or fuzzy) and are most often represented by points, lines and polygons. Continuous spatial phenomenon do not have clearly defined boundaries and thus are best represented as surfaces. Continuous spatial phenomenon are most often represented by a matrix, more commonly referred to as a raster.

Graphic: The Graphic class describes the graphic devices uses to represent and store the abstract representation of spatial phenomenon. Graphics are the drawing primitives available to a cartographer when creating a map. The primitives are point, line and polygon (vector) and matrix (raster). These primitives are mirrored in the map ontology as subclasses of Graphic.

Graphics do not hold any visualization information other than basic shape. The basic shape may be constructed from coordinate pairs, by the hand of a cartographer or both. The pairing of the graphic and visual variables provide the vehicle for visualization on the map body by the cartographer.

Visual Variable: The Visual Variable class represents the concepts cartographers use to visualize the graphics abstracting reality on the map body. This class has nine children covering the most accepted visual variables: arrangement, focus, hue, orientation, saturation, shape, size, texture and value. The names of the subclasses coincide with the visual variable names and will not be discussed here as they are considered to be well known.

Attribute Element: The Attribute class contains the descriptive information related to the spatial phenomenon represented by graphics. Four subclasses, Ordinal, Ratio, Nominal and Interval compose the Attribute class. Attribute class may be visualized on the map body through the Label class, which is a subclass of Layout Element. Typically a label will be placed in a way in which the map reader will related the attribute label to the visualized graphic, but this is not a requirement of the Attribute class or Label class.

4. CONCLUSIONS

The development of this map ontology plays an important role in working towards the partial automation of the map production process with expert systems. Formalizing the vocabulary, structure and relationships of a map for use by a computer program will provide a foundation for software developers, cartographers and software engineers to build expert systems upon.

Choosing to represent the complex subject that is a map in an ontology provides many advantages. Two significant advantages are software independence and accessibility. Software

independence means that the knowledge of maps and mapping are represented independently of any software package. While this map ontology was built with an expert system in mind, it was kept general enough to be used by many other technologies in many other situations. Accessibility means that formalizing the concept of a map in a structured, well-defined and widely used language will facilitate shared expansion of the ontology. Changes and improvements by the community will benefit all interested in modeling the map.

5. REFERENCES

Agarway, P. 2005. Ontological considerations in GIScience. *International Journal of Geographical Information Science* 19 (5):501-536.

CyCorp. 2010. OpenCyc. *Overview of OpenCyc*. http://www.cyc.com/cyc/opencyc (last accessed 9 August 2010).

Fernandez, M., A. Gomez-Perez, and N. Juristo. 1997. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI. www.aaai.org.

Giarratano, J. C., and G. D. Riley. 2005. *Expert Systems, Principles and Programming*. Course Technology.

Gruber, T. R. 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* 43:907-928.

Guarino, N. 1997. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, International Summer School, SCIE-9*, ed. M. Pazienza, 139-170. Franscati, Italy.

Iosifescu-Enescu, I., and L. Hurni. 2007. Towards Cartographic Ontologies or "How Computers Learn Cartography".

Mattmiller, B. 2006. Technology helps foster 'democratization of cartography' (Sept. 20, 2006). *University of Wisconsin-Madison News*. http://www.news.wisc.edu/12916 (last accessed 12 March 2010).

Monmonier, M. S. 1984. Geographic Information and Cartography. *Progress in Human Geography* 8:381-391.

Muller, J. C. 1983. Ignorance graphique ou cartographie de l'ignorance. *Cartographica* 20 (17-30).

Uschold, M., and M. Gruninger. 1996. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review* 11 (2).

W3C. 2004. OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref/ (last accessed 9 September 2010).