Ontology-driven geographic information systems — **Source link** ⬈

Frederico T. Fonseca, Max J. Egenhofer

Institutions: National Center for Geographic Information and Analysis, University of Maine

Related papers:

- A translation approach to portable ontology specifications

- Using Ontologies for Integrated Geographic Information Systems

- Overcoming the semantic and other barriers to GIS interoperability

- Ontology and Geographic Kinds

- Toward principles for the design of ontologies used for knowledge sharing

Electronic Theses and Dissertations   Fogler Library

5-2001

Ontology-Driven Geographic Information Systems

Frederico Torres Fonseca

ONTOLOGY-DRIVEN GEOGRAPHIC INFORMATION SYSTEMS

By

Frederico Torres Fonseca

B.S. Federal University of Minas Gerais - Brazil, 1977

B.E. Catholic University of Minas Gerais - Brazil, 1978

M.S. Joao Pinheiro Foundation - Brazil, 1997

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

May, 2001

Advisory Committee:

Max J. Egenhofer, Professor of Spatial Information Science and Engineering, Advisor

Peggy Agouris, Assistant Professor of Spatial Information Science and Engineering

Claudia M. Bauzer Medeiros, Professor of Computer Science, IC-UNICAMP, Brazil

M. Kate Beard-Tisdale, Professor of Spatial Information Science and Engineering

David M. Mark, Professor of Geography, State University of New York, Buffalo

ONTOLOGY-DRIVEN GEOGRAPHIC INFORMATION

SYSTEMS

By

Frederico Torres Fonseca

Thesis Advisor: Dr. Max J. Egenhofer

An Abstract of the Thesis Presented
in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
(in Spatial Information Science and Engineering)
May, 2001

Information integration is the combination of different types of information in a framework so that it can be queried, retrieved, and manipulated. Integration of geographic data has gained in importance because of the new possibilities arising from the interconnected world and the increasing availability of geographic information. Many times the need for information is so pressing that it does not matter if some details are lost, as long as integration is achieved. To integrate information across computerized information systems it is necessary first to have explicit formalizations of the mental concepts that people have about the real world. Furthermore, these concepts need to be grouped by communities in order to capture the basic agreements that exist within different communities. The explicit formalization of the mental models within a community is an ontology.

This thesis introduces a framework for the integration of geographic information. We use ontologies as the foundation of this framework. By integrating ontologies that are linked to sources of geographic information we allow for the integration of geographic information based primarily on its meaning. Since the

integration may occurs across different levels, we also create the basic mechanisms for enabling integration across different levels of detail. The use of an ontology, translated into an active, information-system component, leads Ontology-Driven Geographic Information Systems.

The results of this thesis show that a model that incorporates hierarchies and roles has the potential to integrate more information than models that do not incorporate these concepts. We developed a methodology to evaluate the influence of the use of roles and of hierarchical structures for representing ontologies on the potential for information integration. The use of a hierarchical structure increases the potential for information integration. The use of roles also improves the potential for information integration, although to a much lesser extent than did the use of hierarchies. The combined effect of roles and hierarchies had a more positive effect in the potential for information integration than the use of roles alone or hierarchies alone. These three combinations (hierarchies, roles, roles and hiearchies) gave better results than the results using neither roles nor hierarchies.

Acknowledgments

moments of his thesis and my thesis, good and bad, my wife Dayse and my daughter Isabela for sharing their lives with me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Information integration is the combination of different types of information in a framework so that it can be queried, retrieved, and manipulated. The specific case of integration of geographic information is the main topic of this thesis. This integration is usually done through an interface that acts as the integrator of information originating from different places.

Integration of geographic information has gained in importance because of the new possibilities arising from the interconnected world and the increasing availability of geographic information. This new information originates from new spatial information systems and also from new and sophisticated data collection technologies. Now information integration is turning into a science (Wiederhold 1999), and it is necessary to find innovative ways to make sense of the huge amount of information available today.

Many times the need for information is so demanding that it does not matter if some details are lost, as long as integration is achieved. For example, frequently sufficient information exists to solve a problem, but integration is difficult to achieve in a meaningful way, because the available information was collected by different agents and with diverse purposes. Events such as the wild fires in and around Los Alamos, New Mexico during the summer of 2000 require a dynamic integration of geographic information. In such a case, a user may be interested in bodies of water that can be used to support the fire extinguishing efforts. In an emergency, the user is not interested in how the information is stored or which data model is being used, but in the value of the information itself, in the meaning of the information. A user wants to know simply and directly "where can I get water; fast?"

For the user in question it does not matter if the information is stored in ArcInfo or in GRASS, two popular GIS software packages. The availability of a growing number of software packages and the ensuing variety of internal data models has created a demand for mechanisms that allow the exchange of geographic information stored in different geographic databases. Early attempts to obtain integration of different GISs involved the direct translation of geographic data from one vendor format into another. A variation of this practice is the use of a standard file format. These formats can lead to information loss, as is often the case with the popular CAD-based format DXF. Alternatives that avoid this problem are also available, but are usually more complex and include the Spatial Data Transfer Standard (SDTS) (USGS 1998) and the Spatial Archive and Interchange Format (SAIF) (Sondheim *et al.* 1999). Although standards for data exchange are necessary and useful for the transfer of large amounts of data, they lack the capability of also transferring the meaning associated with the piece of information when it was first created.

A common format alone is not enough to provide information integration based on meaning (Mark 1993). A growing interest in the development of a common data model led to new lines of research in geographic information integration. One of the largest initiatives following this line of research is the OpenGIS™ Consortium (McKee and Buehler 1996). This association of software developers, government agencies, and systems integrators aims at defining a set of requirements, standards, and specifications to support GIS interoperability. The development of the OpenGIS data model deals primarily with representations of geographic information. New approaches are needed to step up to a higher level of abstraction where the more valuable information about the meaning of the data can be handled. Neither a standard data format nor a common data model allows for the transfer of the meaning of information. The more complex issue of *what* is represented instead of *how* it is represented needs to be addressed. For instance, the user looking for water in New Mexico can obtain this information from the files of the Environmental Protection Agency or from information stored by the New Mexico Parks and Recreation Department. The important thing here is that these two agencies share the same

concept of what a body of water *is*. An active agent that uses this concept can actively look for this information, retrieve it, and make it available for the user.

For integration to be efficient and to deliver the kind of information that the user is expecting, it is necessary to have an agreement on the meaning of words. In a broader scope, it is necessary to reach an agreement about the meaning of the entities of the geographic world. In this thesis the term *semantics* is used to refer to the basic meaning of these entities. These entities are parts of a mental model that represents concepts of the real world, or more specifically, of the geographic world. A concept such as *body of water* carries with it a definition and the mental image that people have of it.

What kinds of agreement can be reached among people? The question whether it is possible to reach such an agreement among all humankind regarding the basic entities of the world belongs to the realm of philosophy and is not part of this investigation. We argue in this thesis that small agreements can be made within small communities. Later, these agreements can be expanded to reach larger communities. When this larger agreement occurs, part of the original meaning is lost, or at least some level of detail is lost. For instance, inside a community of biology scholars, a specific body of water in the state of New Mexico can be a lake that serves as the habitat for a specific species and, therefore, it can have a special concept or name to refer to it. Nonetheless, it is still a body of water, and when a biologist is working at a more general level it is considered as a body of water and not as a lake. At this higher level it is more likely that this real-world entity–body of water–can find a match with the same concept in another community. So the biologist and some member of another community can exchange information about bodies of water. The information will be more general than when the body of water is seen as the habitat of a specific fish species.

For this kind of integration of information to happen among computerized information systems it is necessary first to have explicit formalizations of the mental concepts that people have about the real world. Furthermore, these concepts need to be

grouped by communities representing the basic agreements that exist within each community. Once these mental models are explicitly formalized, mechanisms must be created for generalizing a specific type of lake into a body of water or for adding sufficient specification to the concept of body of water that it becomes a specific lake. People perform such operations in their minds all the time. The requirement to formalize them comes from the need to have these operations available as computer implementations.

Such an explicit formalization of our mental models is usually called an ontology. The basic description of the real things in the world, the description of what would be the truth, is called *Ontology* (with an upper-case O). The result of making explicit the agreement within communities is what the Artificial Intelligence community calls *ontology* (with a lower-case o). Therefore, there is only one Ontology, but many ontologies. This thesis uses the second option, because the goal is to integrate the information that represents the view of diverse communities, each one with its own ontology. We argue that these different views, expressed as ontologies, can be integrated across different levels of detail.

In this thesis we introduce a framework for the integration of geographic information. Ontologies are used as the foundation of this framework. By integrating ontologies that are linked to sources of geographic information we create a mechanism that allows geographic information to be integrated based primarily on its meaning. Since the integration may occur across different levels, as in the case of a body of water and a lake, we also create the basic mechanisms for changes of levels of detail. The use of an ontology, translated into an active, information-system component, leads to Ontology-Driven Information Systems (ODIS) (Guarino 1998) and, in the specific case of GIS, it leads to what we call Ontology-Driven Geographic Information Systems (ODGIS) (Fonseca and Egenhofer 1999).

1.1   Representing Ontologies: Hierarchies and Roles

The example of the biologist's view of a lake presents a series of questions. First, related to semantics, we can ask, "what does a body of water, a lake, a habitat mean?"

or "how many communities, or better, which communities, share the same concept of body of water?"

The communities that offer the information to share (i.e, the information producers) or the communities that want access to information (i.e., the consumers of information) each have an ontology. Each of these ontologies may be subdivided into smaller ontologies. The level of detail of the ontologies is related to the level of detail of the geographic information. Information should also be integrated at different levels of detail. Therefore, two of the main questions of this thesis are "how can these ontologies be combined, leading to information integration?" And, "what are the mechanisms for change of levels inside ontologies?"

The goal of this thesis is to find a mechanism for integrating ontologies and, consequently, for integrating geographic information. This mechanism should provide a way to navigate at different levels in the ontology structure, because in order to answer user queries it is necessary to combine information at different levels of detail and consolidate information on a specific level.

Since ontologies are the foundation of the solution created here for geographic information integration, how they are represented becomes a key factor in the solution. One common solution is to use hierarchies to represent ontologies. Hierarchies are also considered a good tool for representing geographic data models (Car and Frank 1994). Besides being similar to the way we organize the mental models of the world in our minds (Langacker 1987), hierarchies also allow for two important mechanisms in information integration: generalization and specialization. Many times it is necessary to omit details of information in order to obtain a bigger picture of the situation. Other times it is mandatory to do so, because part of the information is only available at a low-level of detail. For instance, if a user wants to see bodies of water and lakes together, and manipulate them, it is necessary to generalize lake to body of water so that it can be handled together with bodies of water. Another solution would be to specialize bodies of water by adding more specific information. Hierarchies can also enable the sharing and reuse of knowledge. We can consider ontologies as repositories

of knowledge, because they represent how a specific community understands part of the world. Using a hierarchical representation for ontologies enables us to reuse knowledge, because every time a new and more detailed entity is created from an existing one it is necessary to *add* knowledge to previous existing knowledge. When we specify an entity lake in an ontology, we can create it as a specialization of body of water. In doing so we are using the knowledge of specialists who have early specified what "body of water" means. The ramifications of reusing knowledge are great and can improve systems specification by helping to avoid errors and misunderstandings. Therefore, we choose to use hierarchies as the basic structure for representing ontologies of the geographic world.

The choice of hierarchies as the representation of the ontologies leaves us with a new problem, however. Many geographic objects are not static: they change over time. In addition, people view the same geographic phenomenon with different eyes. The biologist, for instance, looks at the lake as the habitat of a fish species. Nonetheless, it is still a lake. For a Parks and Recreation Department the same entity is a lake, but it is also a place for leisure activities. Or legislation might be passed that considers the same lake as a protected area. For instance, the biologist's lake can be created by inheriting from a specification of lake in a hydrology ontology and from a previous specification of habitat in an environmental ontology. One of the solutions for this problem is the use of multiple inheritance. In multiple inheritance a new entity can be created from more than one entity. Multiple inheritance has drawbacks, however. Any system that uses multiple inheritance must solve problems such as name clashes, that is, when features inherited from different classes have the same name (Meyer 1988). Furthermore, the implementation and use of multiple inheritance is non-trivial (Tempero and Biddle 1998). We chose to use objects with *roles* to represent the diverse character of the geographic entities and to avoid the problems of multiple inheritance. This way an entity *is* something, but can also play different roles. A lake is always a lake, but it can play the role of a fish habitat or a role of a reference point. Roles allow not only for the representation of multiple views of the same phenomenon, but also for the representation of changes in time. The same building that was a factory in the past must be remodeled to function as an office building. So it

is always a building, but a building playing different roles over time. In our framework, roles are the bridge between different levels of detail in an ontology structure and for networking ontologies of different domains.

1.2   Goal and Hypothesis

This thesis introduces a framework based on ontologies to integrate geographic information. One of the main characteristics of such a framework is its support for information integration. The integration is accomplished through the integration of ontologies. The entities in the ontologies are linked to the information sources; therefore, the integration of ontologies leads to integration of associated information. The integration of ontologies and the inherent issues associated with it are among the main problems that drive the development of this thesis. Specifically, we are investigating the following questions:

- What are the components that influence most the amount of geographic information that can be integrated?

- How can the potential for information integration be measured?

    The answer to the first question leads to the development of a framework for geographic information systems based on ontologies. The framework stresses the importance of hierarchies in the representation of models of the geographic world. The framework also makes use of roles. Each entity in an ontology can play many roles. The answer to the second question leads to the development of a method to evaluate the potential for information integration when combining two ontologies. The hypothesis of this thesis is:

    A model that incorporates hierarchies and roles has a potential to integrate more information than models that do not incorporate these concepts.

    In the approach used by this thesis, information is integrated after the integration of ontologies. Therefore, the approach to test the hypothesis is to measure the potential for information integration after combining ontologies. We developed a method to

evaluate the potential for information integration. This evaluation took into account how the use of roles and hierarchies for representing ontologies influenced the potential for information integration.

We conducted a simulation in which two randomly generated ontologies were combined and the resulting potential for information integration was measured. The measurements were made for ontologies that (1) used roles, (2) used roles and hierarchies together, (3) used hierarchies alone, and (4) used no roles and no hierarchies. We found that the hypothesis is supported by the analysis of the simulation of the integration of two ontologies.

1.3   Scope of the Thesis

Goodchild *et al.* (1999b) define GIScience as the systematic study according to scientific principles of the nature and properties of geographic information. GIScience is mainly concerned with three areas, the *individual*, the *system*, and the *society*. This thesis addresses the interface between *individuals* and *systems*. We start with the *individual*, using a person's perception of the geographic world formalized through geo-ontologies. Then we move to computer implementations of ontologies and the associated mechanisms to deal with them. The classes extracted from ontologies can be used to build GIS applications in the *system* area.

This thesis focuses on the creation of mechanisms to be used in the integration of ontologies. Since the ontologies are linked to the information sources, the integration of ontologies will result in the integration of geographic information. We develop a methodology for the development of geographic information systems based on ontologies. Mechanisms that provide changes of level of detail are also explored in this work. A measure of the potential for information integration when combining two ontologies is also developed here.

This thesis does not attempt to create substantive theories of spatial objects and their relations. Our intention is to offer a framework within which such theories can be used to help the integration of geographic information. Throughout this thesis we use

simplified theories that can be part of a more complete ontology of the geographic world. Most of the examples are based on a subset of two ontologies, WordNet (Miller 1995) and SDTS (USGS 1998), which were combined in Rodríguez (2000).

1.4   Major Results

The major result of this thesis is the specification of a framework based on ontologies for the integration of geographic information The framework allows integration of information at different levels of detail. Since there is not a unifying concept of space (Frank 1997) it is necessary to be able to deal with multiple views of the geographic world. Therefore, it is necessary for GIS developers to be able to integrate different ontologies. The solution presented here allows for the integration of ontologies and the integration of information associated with the ontologies. The integration is accomplished through the combination of classes derived from multiple ontologies. In this way it is possible to create geographic entities that are able to represent the complexity of the geographic world.

The possibility of having multiple views of a single geographic object is provided by the use of hierarchies and roles to support the representation of ontologies. Therefore, a geographic object can have more than one description. The support of multiple interpretations of the same geographic area answers the questions regarding different applications over the same region (Gahegan and Flack 1996). This approach also addresses issues regarding manipulations of different levels of detail of the same object by different applications (Hornsby 1999; Fonseca *et al.* 2000).

An experiment with the integration of randomly generated sets of ontologies tested the hypothesis that a model that incorporates hierarchies and roles has a potential to integrate more information than models that do not incorporate these concepts. We evaluated the influence of the number of roles and the hierarchical structure for representing ontologies on the potential for information integration. We observed a strong influence of the number of roles in increasing the potential for information integration. The use of a hierarchical structure also improved the potential for information integration, although to a much lesser extent than did the use of roles.

The combined effect of roles and hierarchies had a more positive effect in the potential for information integration than the use of roles only or hierarchies only. All those three combinations gave better results than the results using neither roles nor hierarchies. These results supported the hypothesis.

1.5   Intended Audience

This thesis is intended for anyone interested in the integration of geographic information, mainly based on its semantic aspect rather than the way data are stored or represented geometrically. People working with the design and development of GIS, and the development of ontology-driven information systems, including researchers interested in geo-ontologies, geographic database design, and geographic object models, will also find material of interest in this thesis. GIScientists concerned with the *individual* and the *system* areas will find this thesis interesting, because it addresses a subject on the interface between these two areas. Computer scientists concerned with implementations of GIS and ontology-driven information systems should also find in this thesis useful material regarding the use of ontologies as components of information systems.

1.6   Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 reviews related work on the use of object orientation and ontologies for the computer representation of conceptualizations of the geographic world. A classification of ontologies according to their level of details is presented. The use of ontologies for information integration is also reviewed. Two implementations of information systems that use ontologies are shown.

Chapter 3 introduces a multiple-ontology approach to geographic information integration. The different kinds of ontology–phenomenological domain ontology and application domain ontology–are introduced. The chapter also discusses vertical and

horizontal navigation inside the framework. The operations of inheritance, inclusion, and role extraction that are used for vertical and horizontal navigation are presented.

Chapter 4 describes a methodology for creating the framework focusing on the aspects of knowledge generation and knowledge use. Then it shows how the ontologies are specified by the geospatial communities. It presents how the knowledge generated in the first phase of the system can be used to develop GIS applications. The mechanism that allows a piece of information to change its level of detail is presented. The different levels of detail of information and their relation to different levels of ontologies are discussed here.

Chapter 5 discusses ontology integration and introduces the concepts of high-level and low-level integration. Also presented in this chapter is a measure of the potential for information integration when combining two ontologies. Two experiments and the results supporting the hypothesis are described. The chapter also concludes that the number of roles has a strong influence in increasing the potential for information integration.

Chapter 6 discusses implementation issues and describes how the main components can be implemented. The chapter analyzes the implementation options for the main components of the framework. The use of Java as an implementation language is discussed. The development of an ontology editor was suggested. The ontology browser is presented. A query for three different entities in an ontology is shown and the results are discussed.

Chapter 7 presents conclusions and future work. The chapter presents the main contributions of the framework for the integration of geographic information and a summary of the work. The methodology for evaluating the potential for information integration when two ontologies are combined is reviewed. The effects of using roles and hierarchies in the potential of geographic information that can be integrated are discussed. Future research regarding further development of the framework is discussed. New problems in ontology integration, geographic information retrieval on

the web, ontology specification, ontology of actions, and ontology of images are suggested as themes for future research.

Chapter 2

Objects and Ontologies for GIS Integration

Research on integration of databases can be traced back to the mid 1980s (Batini *et al.* 1986), and today it is widespread among the GIS community (Worboys and Deen 1991; Kashyap and Sheth 1996; Bishr 1997; Bishr 1998; Mena *et al.* 1998; Gahegan 1999; Goodchild *et al.* 1999a; Harvey 1999). The complexity and richness of geographic information and the difficulty of its modeling raise specific issues for GIS interoperability, such as the integration of different models of geographic entities (i.e., objects and fields) and different computer representation of these entities (i.e., raster and vector).

The literature shows many proposals for the integration of information, ranging from federated databases with schema integration (Sheth and Larson 1990) and the use of object orientation (Kent 1993; Papakonstantinou *et al.* 1995), to mediators (Wiederhold 1991) and ontologies (Wiederhold 1994; Guarino 1998). The new generation of information systems should be able to handle semantic heterogeneity in making use of the amount of information available with the arrival of the Internet and distributed computing (Sheth 1999). The semantics of information integration is getting more attention from the research community (Worboys and Deen 1991; Kuhn 1994; Kashyap and Sheth 1996; Bishr 1997; Câmara *et al.* 1999; Gahegan 1999; Harvey 1999; Sheth 1999; Rodríguez 2000). The support and use of multiple ontologies should be a basic feature of modern information systems if they want to support semantics in the integration of information. Ontologies can capture the semantics of information, can also be represented in a formal language, and can be used to store the related metadata enabling this way a semantic approach to information integration.

We argue that sophisticated structures, such as ontologies, are good candidates for abstracting and modeling geographic information. Our solution is based on a semantic approach using the concept of geographic entities (Nunes 1991). The next section shows the importance of the use of an object model to model the geographic world, followed by a discussion of GIS interoperability and the use of ontologies to achieve it. Then we review system architectures for integrated GIS and ontology-driven systems. The last section of this chapter presents a summary of the chapter.

2.1   An Object View of the World

The use of the object data model as the basic conceptualization of space has been discussed before in the literature. The issue of defining geographic space is actually the issue of defining and studying the geographic objects, their attributes, and relationships (Nunes 1991). The object view of the spatial world (Egenhofer and Frank 1992) avoids problems such as the horizontal and vertical partitioning of data (Kuhn 1991), although objects can provide both, if necessary. Furthermore, an object representation of the geographic world offers many views of a geographic entity. Objects are also useful in zooming operations, because when we get closer to a scene, instead of seeing enlarged objects we see different kinds of objects (Tanaka and Ichikawa 1988; Volta and Egenhofer 1993; Timpf and Frank 1997). These operations are performed through aggregation as in the case of a house constituted by walls and a roof, or a block formed by land parcels (Kuhn 1991).

We model geographic phenomena using an object-oriented approach. This approach should not be mistaken by the conceptualization for the representation of the geographic world. The most accepted models for representation are the *object* and *field* models (Couclelis 1992; Goodchild 1992). The *object model* represents the world as a surface occupied by discrete, identifiable entities with a geometrical representation and descriptive attributes. These objects are not necessarily related to a specific geographic phenomenon and they can be constructed features, such as roads and buildings. The *field model* views geographic reality as a set of spatial distributions over geographic space. Climate and vegetation cover are typical

examples of geographic phenomena modeled as fields. Although this simple dichotomy has been subject to criticism (Burrough and Frank 1996), it has proven to be a useful frame of reference and has been adopted, with some variations, in the design of the current generation of GIS technology (Câmara *et al.* 1996). We accept this model and use it for the representation of geographic entities.

A *class* is the extension of the concept of an abstract type, a structure that represents a single entity, describing both its information content and its behavior. A class defines the structure and the set of operations that are common to a group of objects (Meyer 1988). An *instance,* or *object*, represents an individual occurrence of a certain class. While the class is the type definition, an instance is the data structure represented in the memory of a computer and manipulated by a software system. In this thesis, the terms *object* and *instance* are used interchangeably.

An object functions as a complex data structure that is capable of storing all of its data, along with information about the necessary procedures to create, destroy, and manipulate itself. In an object-oriented GIS, for instance, the separation of spatial and non-spatial attributes is avoided because everything is stored together.

The ability to hide from the user the internal structure of an object is called *encapsulation*. With encapsulation it is possible to manipulate the object's data only by using a set of predefined functions. This approach ensures data independence: the internal implementations of the data structure used by the object can change without influencing what the user perceives.

One of the most important concepts in object-oriented systems is *inheritance*. Inheritance is a classification mechanism in which a class can be the subclass of another (i.e., it incorporates the other's features in addition to its own). Features can be attributes, functions or rules. A subclass is called a descendant. A superclass is any class that is up in the direct hierarchy. When a given class inherits directly from only one superclass, it is called *single inheritance*; when a class inherits from more than one immediate superclass, it is called *multiple inheritance* (Cardelli 1984). Multiple inheritance is a controversial concept, with benefits and drawbacks. For instance, any

system that uses multiple inheritance must provide an adequate solution to problems such as name clashes (i.e., when features inherited from different classes have the same name). Although the implementation and use of multiple inheritance is non-trivial (Tempero and Biddle 1998), its use in geographic data modeling is essential (Egenhofer and Frank 1992). In order to avoid the problems of multiple inheritance and at the same time represent the diverse character of the geographic entities we introduce the concept of *roles*.

2.2   Objects with Roles

An object is something–it has an identity (Hornsby 1999)–but it can play different roles. Usually the notion of role is linked with change in time. An object is only one thing but it can play different roles during its lifetime. The use of roles in object orientation is reviewed in detail by Pernici (1990), Albano *et al.* (1993), Wong (1997), and Steimann (2000). The use of roles in the specification of ontologies is discussed in Guarino (2000a). The concept of role as interfaces as we use in the implementation of this thesis is reviewed in Steimann (2001).

One of the most common use of roles is to represent changes in an object during its lifetime. The typical example is of a person that plays the roles of a student, a parent, and a member of a club. In this thesis roles also help to express different points of view of the same phenomenon. One community may see a certain phenomenon X and consider that X is a occurrence of an entity A. Another community may classify the same phenomenon X as being B. For this second community, B may also play a role of A.

The main objective of using roles in this thesis is to employ them as a tool to connect different ontologies. Therefore we use here a more unrestrained definition of roles than other authors (Guarino and Welty 2000a) who argue that roles should have their own hierarchy and can only subsume or be subsumed by another role. Some authors consider that an object can play a role only if the role is a subtype (Bock and Odell 1998) or a supertype (Halbert and O'Brien 1987) of the object. This point of

view is not adopted here, because for us a role is an entity. Each community has a right to its own point of view and information must be integrated on that basis, hence an use of a flexible specification of role. A more rigid specification would require, for instance, a habitat to be a subclass of a geographical region. As a consequence, in a biologist's ontology, a habitat would not be an entity but only a role. Using a more flexible specification of role we can allow a habitat to be an entity. In this specific point of view, a habitat has an identity and all the attributes that characterize an entity as being distinct from other entities. In our framework every role is an entity. An entity plays roles that are entities in other ontologies.

For instance, for a biologist a habitat can play a role of a lake or a role of woods near the lake. Some authors would argue that habitat is only a role and should be always played by a geographic location. We do not agree with this argument. In our framework a habitat is an entity in a biologist's ontology. He/she can work with the entity habitat having all the characteristics of a lake. He can also use a role of lake. He/she can reuse the entity lake avoiding to redefine all of its properties again. Using lake as a role instead of as a superclass gives the biologist more flexibility. He/she can have habitat inherit from a more related entity in his/her biologic point of view, thus avoiding too strong a geographic point of view. Another reason for using lake as a role is for obtaining metadata and data from other sources.

A role can be viewed in different ways (Steimann 2000). First, a role is viewed as a named relationship. This point of view stresses that roles exist only within some particular context. Second, a role is viewed a specialization or a generalization. The problem with this point of view is that it contradicts Guarino's (1992) and mixes the dynamic nature of the role concept with the rigid properties of a type hierarchy. Finally, roles can be represented as adjunct instances. In this point of view, roles are considered totally dependent on the instances that play them and do not carry their own identity. The object and its roles form an aggregate.

We choose here to use roles as adjunct instances for two main reasons. First, we consider roles and types to be parts of separate and independent hierarchies. Second,

the use of adjunct instances is more in accordance with our mechanism to extract roles and with our implementation based on delegation. The extraction operation is one of the features that roles can have.

The extraction of roles and the resulting generation of a new instance of a class can be classified by what is called in the literature as *object migration* or *dynamic reclassification* (Su 1991; Mendelzon *et al.* 1994). The term migration is used to model the change from one role to another in systems in which class membership is the main mechanism for assigning roles. Dynamic reclassification by role-based systems enable objects to dynamically change types and classes membership. This concept can be extended into *multiple classification*, (allowing an object to be an instance of multiple classes), *dynamic reclassification*, (allowing an object to gain and lose class memberships throughout the object's lifetime), and *dynamic restructuring*, (allowing an object's structure to change dynamically throughout the object's lifetime) (Kuno and Rundensteiner 1996).

2.3   GIS Interoperability

Despite initiatives such as SDTS, SAIF, and OpenGIS, the use of data transfer standards as the only worthwhile effort to achieve interoperability is not widely accepted. Since widespread heterogeneity arises naturally from a free market of ideas and products, it is difficult for standards to banish heterogeneity by decree (Elmagarmid and Pu 1990). The use of semantic translators in dynamic approaches is a more powerful solution for interoperability than the current approaches that promote standards (Bishr 1997).

Another important question in GIS interoperability is semantics. Considering the complex issue of the meaning of information and its description, three types of heterogeneity are distinguished (Bishr 1998):

- semantic heterogeneity, in which a fact can have more than one description or interpretation;

- schematic heterogeneity, in which the same object in the real word is represented using different concepts in a database; and

- syntactic heterogeneity, in which the databases use different paradigms.

A set of rules and constraints should be attached to the object class definitions in order to overcome semantic heterogeneity, which should be solved before schematic and syntactic heterogeneity (Bishr 1998).

The idea of a virtual space where different conceptualizations would meet is also discussed in the literature. The Virtual DataBase system (VDB) is an architecture to integrate and retrieve information from multiple component systems, distributing the processing load through the global front end and the components. VDB is based on an object-oriented model and uses the schema integration approach (Abel *et al.* 1998). The Virtual Data Set (VDS) uses a well-defined canonical interface to access multiple spatial databases. VDS corresponds to a protocol between the data consumer and the data producer. VDS is also based on the object orientation paradigm (Vckovski 1997).

The concept of object orientation to provide interoperability can be used either in the implementation or in the modeling phase of system development. The ability to represent complex data structures and behavioral specifications is seen as a reason for using object technology in interoperation (Soley and Kent 1995). Object orientation has some features that are useful to enhance information compatibility, such as the use of object identity to link different sources and reconciliation of different levels of abstraction through subtyping (Kent 1993). Clients prefer to receive information in an object-oriented format when integrating multiple heterogeneous sources, because objects enable aggregation of information into meaningful units. These units can have hierarchical linkages to other classes and so can provide a valid model even for a complex world (Papakonstantinou *et al.* 1995; Wiederhold 1998). Other lines of research in interoperability consider different solutions such as the use of ontologies as the common point among diverse user communities (Wiederhold 1994). The use of ontologies to enable interoperation is the theme of the next section.

2.4 Ontology and Interoperation

The foundation of ODGIS is the willingness of users to share information. The reasons to do so can be economic or regulatory. Reusing information can dramatically decrease the costs of developing a GIS project and can also be a positive factor in the success of a project (Huxhold 1991). Since it is difficult to lower these costs it is better to focus research on sharing the knowledge already acquired. Sharing is a way to build qualitatively larger knowledge-based systems, because we can rely on previous labor and experience (Neches *et al.* 1991). Many high-level government institutions recommend the use of mechanisms that enhance the possibility of information sharing (Arctur *et al.* 1998).

For interoperability to take place, an agreement on the terminology in the shared area must occur through the definition of an ontology for each domain (Wiederhold 1994). Ontologies are crucial for knowledge interoperation, and they can serve as the embodiment of a consensus reached by a professional community (Farquhar *et al.* 1996). Sharing the same ontology is a pre-condition to information sharing and integration. There should be an ontological commitment revealing the agreement between the generic user querying the database and the database administrator that made the information available (Kashyap and Sheth 1996). An alternative to an explicit ontological commitment is the semantic approach. One solution is the derivation of a global schema to overcome the absence of a common shared ontology through the use of clustering techniques. This way the solution of semantic heterogeneity is done through description logic (Bergamaschi *et al.* 1998). Another semantic approach is a similarity assessment among ontologies using a feature-matching process and semantic distance calculations (Rodríguez *et al.* 1999). In ODGIS, the agreement is expressed through the use of elected ontologies that are used to derive new ontologies, from which the software components are derived.

Who are the producers and users of the ontologies used in ontology-driven information systems? We can group the users of geographic information into geospatial information communities (GIC) according to their conceptualizations of the world. The definition of a GIC should not be restricted to users that share the same

data model. Hence we can use the definition of a GIC as a group of users that share an ontology (Bishr 1997). In the solution presented here, we allow the GIC to commit to several ontologies. The users have means to share information through the use of common classes derived from ontologies.

Semantic translators are one of the means to provide interoperability among and within GICs. Semantic translators, also called mediators (Wiederhold 1991), use a common ontology library as a measure of semantic similarity. Dynamic approaches for information sharing, as provided by semantic translators, are more powerful than the current approaches that promote standards (Bishr 1997). Mediation is also proposed as the principal means to resolve semantic heterogeneity through an incremental domain approach that brings domains together when needed. Mediators look for geographic information and translate it into a format understandable by the end user. The mediators are pieces of software with embedded knowledge. Experts build the mediators by putting their knowledge into them and keeping them up to date (Wiederhold 1994).

2.5 Ontology Levels

In the ODGIS architecture there are different levels of ontologies. Accordingly, there are also different levels of information detail. There is a distinction is between coarse and fine-grained ontologies. A coarse ontology consists of a minimal number of axioms and is intended to be shared by users that already agree on a conceptualization of the world. A fine-grained ontology needs a very expressive language and has a large number of axioms. Coarse ontologies are more likely to be shareable and should be used on-line to support the system's functionality. On the other hand, fine-grained ontologies should be used off-line, because they are accessed eventually for reference purposes. Our solution allows the user to incrementally go from coarse to fine-grained ontologies on-line, thus eliminating the division between on-line and off-line ontologies (Guarino 1998).

In this thesis we use the term *low-level* ontologies for fine ontologies and they represent very detailed information and *high-level* ontologies for coarse ontologies and

they represent more general information. Thus, if a user is browsing high-level ontologies he or she should expect to find less detailed information. We propose that the creation of more detailed ontologies should be based on the high-level ontologies, such that each new ontology level incorporates the knowledge present in the higher level. These new ontologies are more detailed, because they refine general descriptions of the level from which they inherit.

Ontologies are classified according to their dependence on a specific task or point of view (Guarino 1997):

- *Top-level ontologies* describe very general concepts. In ODGIS a top-level ontology describes a general concept of space. For instance, a theory describing parts and wholes, and their relation to topology, called mereotopology (Smith 1995), is at this level.

- *Domain ontologies* describe the vocabulary related to a generic domain, which in ODGIS can be remote sensing or the urban environment.

- *Task ontologies* describe a task or activity, such as image interpretation or noise pollution assessment in ODGIS.

- *Application ontologies* describe concepts depending on both a particular domain and a task, and are usually a specialization of them. In ODGIS these ontologies are created from the combination of high-level ontologies. They represent the user needs regarding a specific application, such as an assessment of lobster abundance in the Gulf of Maine.

Representing geographic entities–either constructed features or natural differentiations on the surface of the earth–is a complex task. They are not merely located in space, they are tied intrinsically to space (Smith and Mark 1998). For instance, boundaries that seem simple can in fact be very complex. An example is the contrast between soil boundaries, which are fuzzy, and land parcels whose boundaries are crisp. Users who are developing an application can make use of the accumulated knowledge of experts that have specified an ontology of boundaries instead of dealing

with these complex issues by themselves. The same is true for ontologies that deal with geometric representations, land parcels, and environmental studies. Users should be able to create new ontologies building on existing ontologies whenever possible. An example of a *backbone taxonomy*, which represents the most important properties in a high-level ontology is given in Figure 2-1 (Guarino and Welty 2000b).
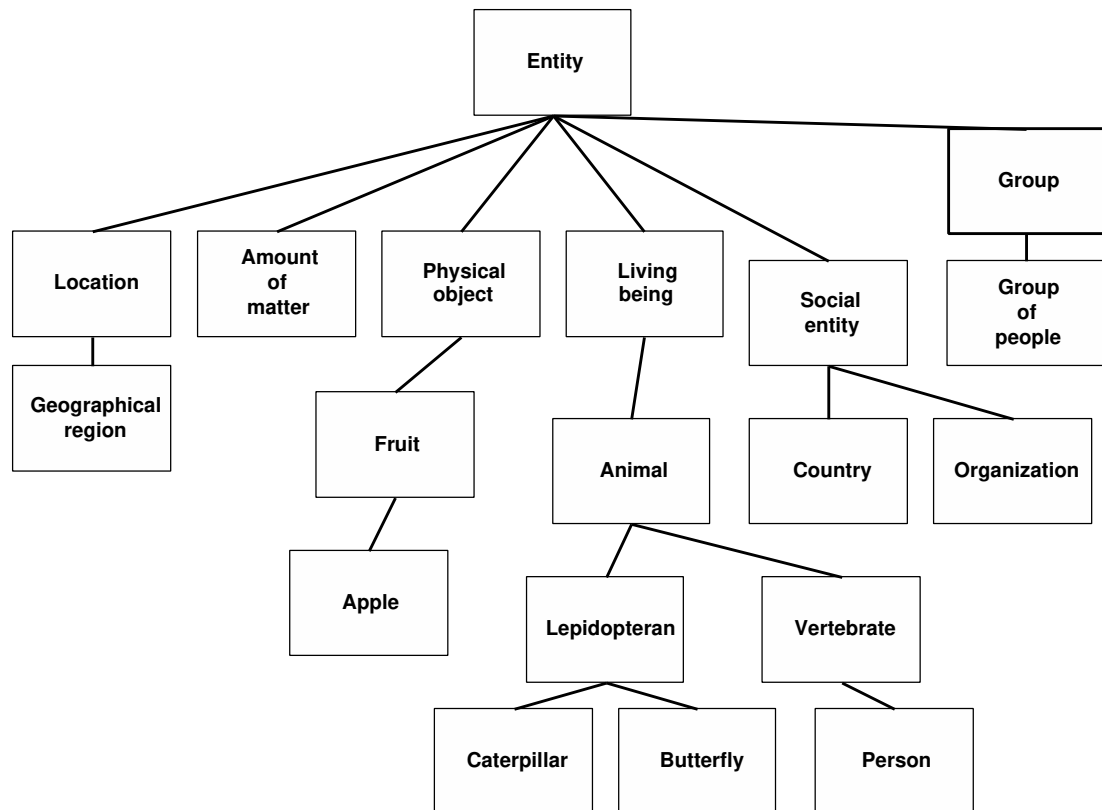
Figure 2-1 A basic taxonomy, from Guarino and Welty (2000).

If a local government is starting a GIS project based on ontologies, we can use a basic urban ontology such as (Huxhold and Levinsohn 1995):

- The geographic coverage of the local government area

- The people within the area

- The buildings and facilities

- The business activities

- The land itself

Instead of defining these four main branches in detail, the users could use the backbone taxonomy introduced before and from it, start their own ontology. A sample result can be seen in Figure 2-2 where the class *People* is derived from the class *Person*, *Business* is derived from *Organization*, and *Land* is derived from *Geographical region*. At the same time, if the urban ontology is general enough, it can be used as the foundation for other local government projects.
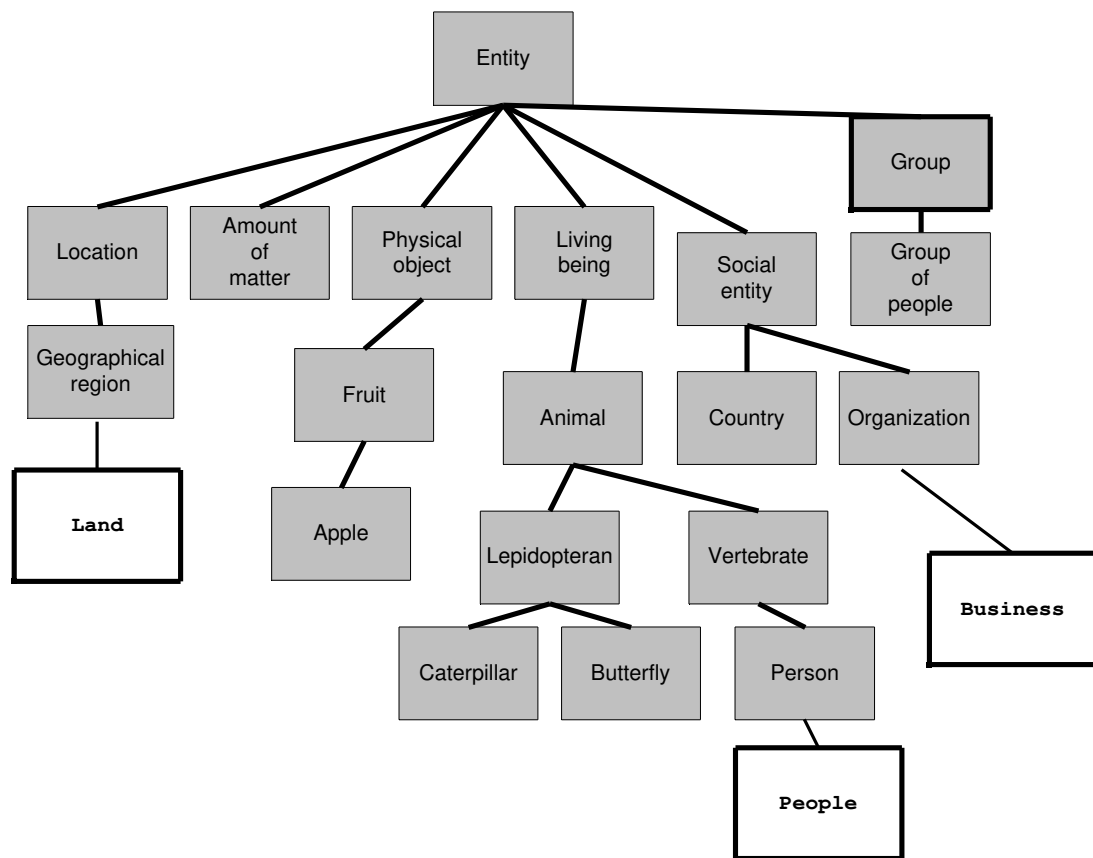
Figure 2-2 Deriving new classes from a high-level ontology.

An application developer can combine classes from diverse ontologies and create new classes that represent user needs. In this way, a class that represents

Building in the urban ontology can be built from *Physical object* in the basic taxonomy. At the same time, *Building* can be seen as a location and can also hold a social entity or an organization. Thus, *Building* can play the roles of *Location* and *Organization* extracted from the urban ontology. So the real class is *Building*, but it plays many roles (Figure 2-3) that together give the class its unique characteristics.
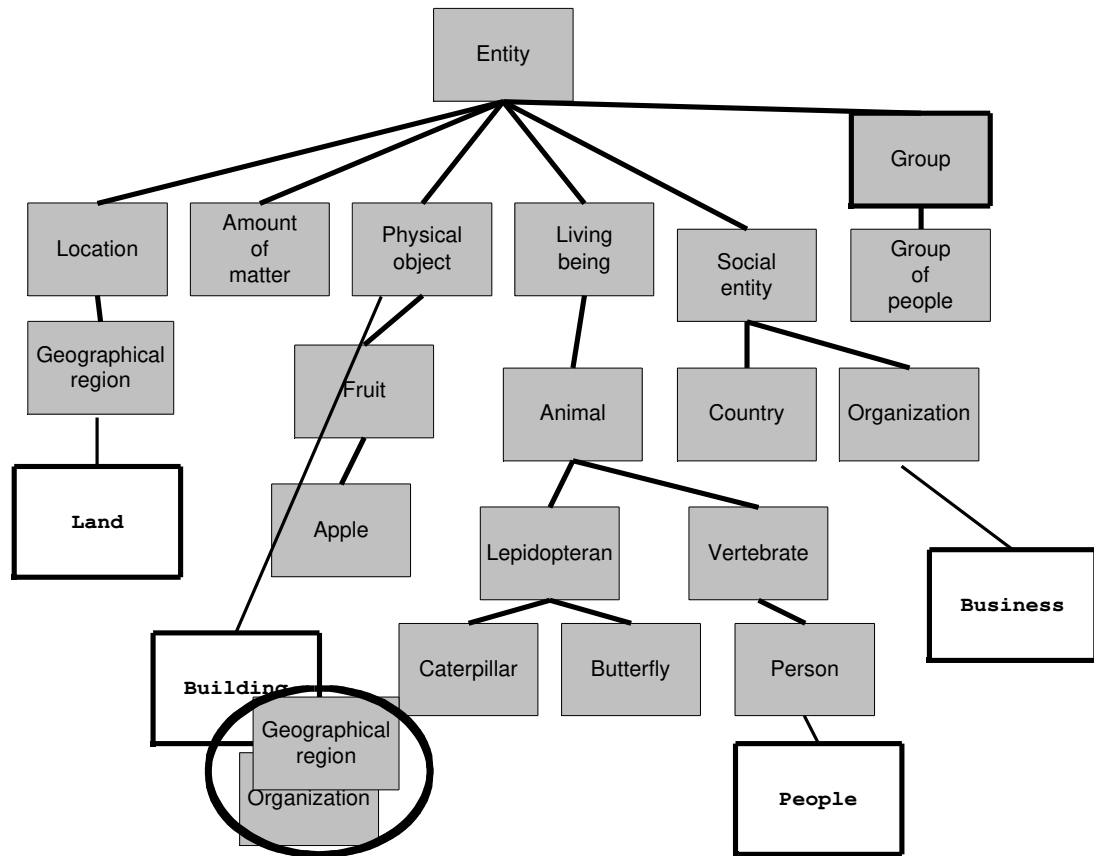
Figure 2-3 A class can play many roles.

2.6 Ontology-Based System Architectures

The new generation of information systems should be able to solve semantic heterogeneity. The support and use of multiple ontologies should be a basic feature of the modern information systems. We review here Ontolingua, a language to specify

ontologies which can be used for these kinds of systems and OBSERVER, an information retrieval system based on ontologies.

2.6.1 Ontolingua

A mechanism to edit, browse, translate, and reuse ontologies is presented in the Ontolingua Server (Farquhar *et al.* 1996), which is based on Ontolingua (Gruber 1992), a language to specify ontologies. The syntax and semantics of Ontolingua definitions are based on the Knowledge Interchange Format (KIF) (Genesereth and Fikes 1992). KIF is a monotonic, first-order predicate calculus with a simple syntax and support for reasoning about relations. The approach used in Ontolingua is to translate ontologies specified in a standard, system-independent form into specific language representations. The Ontolingua Server allows multiple users to collaborate on ontology construction in a shared section. It also accepts queries from remote applications. The Ontolingua translation strategy allows the use of an ontology both in the development and in the production phases of a system. The translation targets can be representations in CORBA interface definition language (IDL) (OMG 1991), Prolog (Clocksin and Mellish 1981), Epikit (Genesereth 1990), or KIF. An excerpt of a graphic representation of an urban ontology is shown in Figure 2-4, an example of the ontology *Simple-Geometry* in Ontolingua is given in Figure 2-5, and a description of the ontology *Quantity-Space* inside the ontology *Simple-Geometry* using the language LISP generated by Ontolingua is given in Figure 2-6.
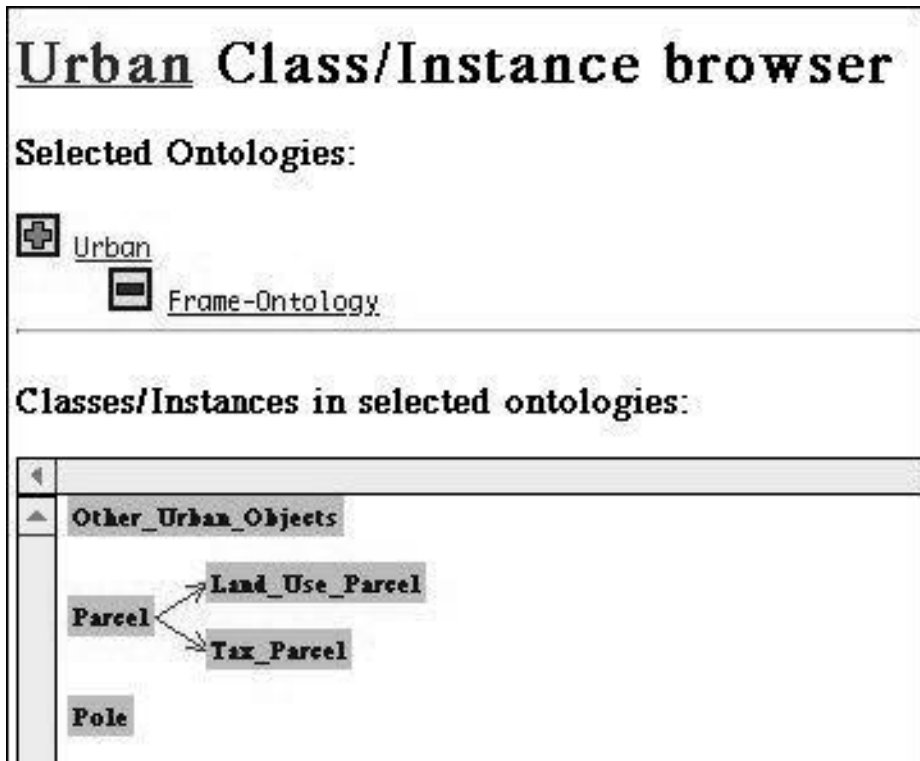
Figure 2-4 A graphic representation of an urban ontology in Ontolingua.

Ontology SIMPLE-GEOMETRY
* **Last modified:** *Tuesday, 2 September 1997*
* **Generality:** High
* **Maturity:** High
* **I/O Syntax:** Case Insensitive
* **Private by default:** No
* **Source code: simple-geometry.lisp**

Ontology documentation:

This ontology attempts to capture basic geometric concepts used in mechanical systems modelling. These concepts include points, frames, position, and orientation but exclude notions of extent.

Summary of Simple-Geometry:

Simple-Geometry includes the following ontologies:

  3d-Tensor-Quantities
  Quantity-Spaces
  Standard-Dimensions

No ontologies include Simple-Geometry.

Class hierarchy (3 classes defined):

    3d-Direction-Cosine
    3d-Frame
    3d-Point

No relations defined.

4 functions defined:

  Distance
  Orientation
  Position
  Simple-Rotation

1 individual defined:

  3d-Length-Space

44 unnamed axioms defined.

No named axioms defined.

Figure 2-5 An example of the ontology *Simple-Geometry* in Ontolingua.

```
(in-package "ONTOLINGUA-USER")
(define-ontology quantity-spaces (physical-quantities)
  "A quantity-space is a set that has the property that a     distance function is
defined for any two elements in the set.  In addition, the range of the distance
function is a subclass of the class of scalar quantities. This ontology defines
the class of quantity-space, and the associated relations POINT-IN,
DISTANCE. It is agnostic about the semantics of the points -- they needn't
be spatial things or of any particular dimensionality."
  :maturity :moderate
  :generality :moderate
  :issues ("Copyright (c) 1994 Greg R. Olsen and Thomas R. Gruber"
          (:see-also "The EngMath paper on line")))
(in-ontology 'quantity-spaces)
(define-class QUANTITY-SPACE (?s)
"A quantity-space is a set that has the property that a distance function is
defined for any two elements in the set.  In addition, the range of the distance
function is a subclass of the class of scalar quantities."
  :iff-def
    (and (set ?s)
          (forall (?x1 ?x2)
                  (=> (and (member ?x1 ?s)
                          (member ?x2 ?s))
                      (exists (?d)
                          (and
                          (= ?d (distance ?x1 ?x2))
                          (scalar-quantity@scalar-quantities ?d)))))))
```

Figure 2-6 The description of the ontology *Quantity-Space* in LISP.

2.6.2   OBSERVER

OBSERVER (Kashyap and Sheth 1996; Mena *et al.* 1996; Mena *et al.* 1998) is an
architecture for query processing in global information systems that supports
interoperation across ontologies. It focuses on information content and semantics, and
employs a loosely-coupled approach to match different vocabularies used to describe
similar information across domains. Instead of integrating pre-existing ontologies,
OBSERVER uses synonym relationships between terms across ontologies. Synonymy,
hyponymy, and hypernymy are semantic relations defined between words and word
senses. Synonymy (*syn* same, *onyma* name) is a symmetric relation between word
forms. Hyponymy (sub-name) and its inverse, hypernymy (super-name), are transitive
relations between sets of synonyms. This semantic relation is usually organized in a

hierarchical structure (Miller 1995). OBSERVER uses hyponymy and hypernymy to translate terms that are not synonymous in different ontologies. It substitutes non-translated terms with the intersection of their immediate parents or the union of the immediate children. The approach used here is to pursue the definition of a method that finds similar entity classes that can link entities in independent databases to achieve information integration (Figure 2-7). Unlike OBSERVER, other solutions do not create new ontologies, but create links between similar entities in distinct ontologies (Rodríguez 2000).
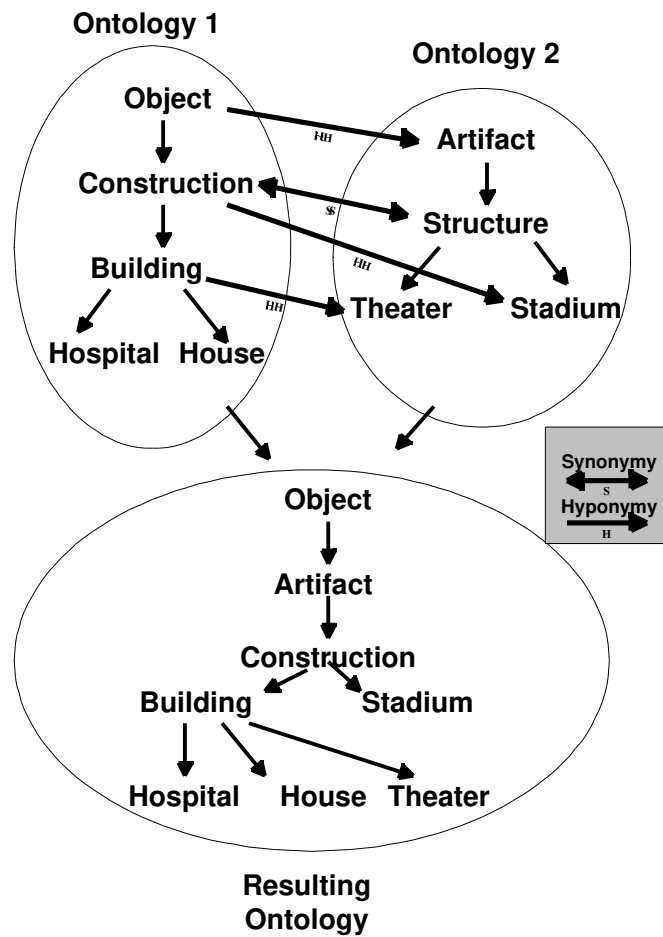
Figure 2-7 Hyponym and synonym relationships, from Rodríguez (2000).

The basic components of OBSERVER are the query processor, the ontology server, and the interontology relationships manager. The user query is based on an

ontology chosen by the user. The query processor matches the terms in the user ontology to the system component ontologies. The ontology server provides information about ontologies using mappings between ontologies and the structures in data repositories. The interontology relationships manager provides the synonym relationships.

2.7   Summary

This chapter reviewed related work on the use of object orientation and ontologies for the computer representation of conceptualizations of the geographic world. The different types of ontologies were presented. The use of ontologies for information integration was also reviewed. Two implementations of information systems that use ontologies were shown.

The next chapter introduces a multiple-ontology approach to geographic information integration. Two kinds of ontology, a phenomenological domain ontology and an application domain ontology, are introduced. The chapter also discusses vertical and horizontal navigation inside the framework and the operations of inheritance, inclusion, and role extraction that are used for navigation are presented.

Chapter 3

A Conceptual Framework for Geographic Information Integration

In order to understand how people see the world and how ultimately the mental conceptualizations of the apprehended geographic features are represented in a computer system we must develop abstraction paradigms. The result of the abstraction process is a general view of the process that goes from the real object to its computer representation. The use of different levels of abstraction allows the development of specific tools for the different types of problems at each level. In this chapter, we introduce a conceptual framework for the understanding and representation of the geographic world. The main components are five universes and the operations that connect them. The concepts presented in this chapter give the foundations for the understanding of ontology-driven geographic information systems.

We introduce the five-universes paradigm, which builds on the four-universes paradigm (Gomes and Velho 1995), by adding new components and explaining some of the concepts from the point of view of the geographic world. Our main contributions to the four-universes paradigm are:

- the addition of the cognitive universe;

- the connection of the cognitive universe to the logical universe; and

- the use of ontologies as the key component of the logical universe.

In the next section we explain how the geographic world can be understood using the five-universes paradigm and which are the operations that interconnect the universes. Then we introduce the multiple-ontology approach for ontology-driven

geographic information systems. This approach enables the reuse of knowledge and a better understanding of the geographic phenomena. Two kinds of ontologies for the geographic world are introduced. One is called *Phenomenological Domain Ontology* and aims at capturing the different dimensions and internal properties of the geographic phenomena. The other type is concerned with description of specific subjects and tasks and is called the *Application Domain Ontology*. The multi-ontology approach leads to bi-directional integration of geographic information. The chapter's summary is the last section.

3.1   An Abstraction Paradigm for the Geographic World

The understanding of the geographic world, with the final objective of having a computer representation, has been the subject of much study in the last decade. In assembling our view of the world we build on previous explanations of how people see and mentally represent the world (Requicha 1980; Couclelis 1992; Goodchild 1992; Gomes and Velho 1995). Each of the five levels in our abstraction model deals with conceptual characteristics of the geographic phenomena of the real world. The first two levels, the physical level and the cognitive level, are only briefly described here. This thesis is concerned mainly with the three last levels, the logical level, the representation level, and the implementation level. Once one level is understood, we are able to face the problems of the next level.

The five universes are the *physical universe*, the *cognitive universe*, the *logical universe*, the *representation universe*, and the *implementation universe* (Figure 3-1). A geographic phenomenon in the real world is captured by the cognitive system of a person and is classified and stored in the human mind. The representation of the real world object in the human cognitive system is done within the cognitive universe. The formalization of the conceptualizations of the world in the human mind gives us explicit formal structures, the ontologies that are part of the logical universe. When we take into account the particularities of the spatial world–for instance, reference systems and conceptualizations such as fields and objects–we are dealing with the representation universe. The shift to the implementation universe is made through the

translation of the components of the representation universe into computer language structures.
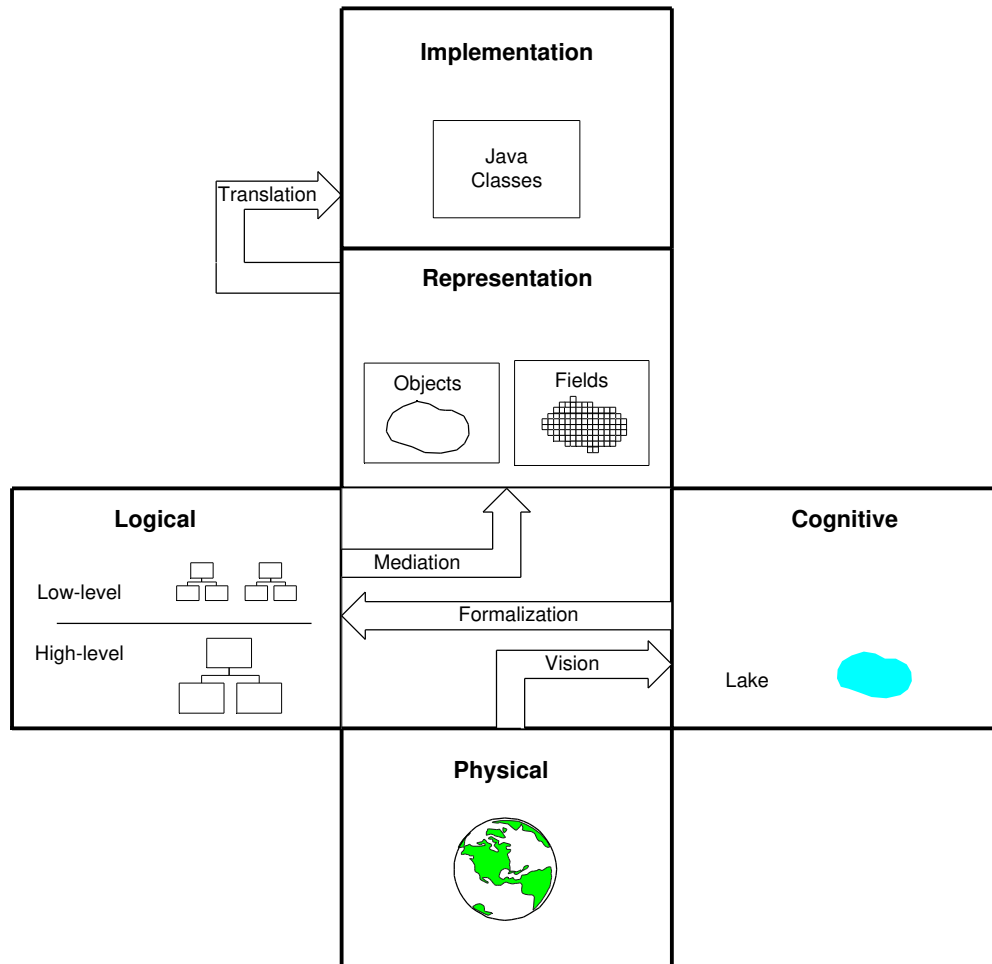
Figure 3-1 The five-universe-paradigm.

The physical universe is the real world with everything that people are capable of perceiving. The real objects are there. Vegetation, rivers, and mountains are part of the real-world phenomena that we are interested in. The process called *vision* (Marr 1982) is the connection between the physical universe and the cognitive universe. Through vision, images that correspond to real world objects are formed inside people's mind. These images in the cognitive universe are representations of the entities in the physical universe. But these images are not merely stored in the mind in a haphazard way; they are organized in a logical framework (Bryant and Tversky

34

1992). When this framework is made explicit using logical methods, we obtain ontologies (Guarino and Welty 2000a). They are the formal representations of the logical schemes of the human mind and they exist in the logical universe.

The logical universe contains two types of ontologies. High-level ontologies contain the more general theories of the world, such as the general concepts of a theory of natural geography. Low-level ontologies are specializations of more general ontologies. They can be detailed descriptions of specific domains and the tasks that deal with these domains. The logical universe is connected to the representation universe by *semantic mediators*.

The representation universe is where a finite symbolic description of the elements in the logical universe is made so that we can apply operations on them. Here the ontologies of objects and fields are defined as the basic conceptualizations of the geographic world. Also here is the place to deal with all the concerns related to how these concepts are captured from the real world and how they are measured. The ontologies present at the representation level and at the logical level can be translated into computer languages, generating classes that belong to the implementation universe.

The implementation universe can include elements, such as algorithms in computer language, vector and raster data structures, and classes in object-oriented languages. In this thesis we deal only with classes that are the result from the translation of entities in ontologies.

3.2   A Multiple-Ontology Approach

Ontologies for the geographic world, the geo-ontologies, should be divided in two types. One type is the *Phenomenological Domain Ontology* (PDO). This ontology captures the different dimensions and internal properties of the geographic phenomena. This specific ontology is distinct and independent from the other type, the *Application Domain Ontology* (ADO). This ontology is concerned with description of specific subjects and tasks that the GI scientists use as a source of information.

Since the PDO is concerned with how the geographic phenomenon can be captured and represented by computer systems, it is located in the representation universe. The ADO is part of the logical universe because, it deals with the description of the phenomenon itself, where it fits in the world, and how it can be best described. The connection between PDO and ADO is made by semantic mediators (Figure 3-2).
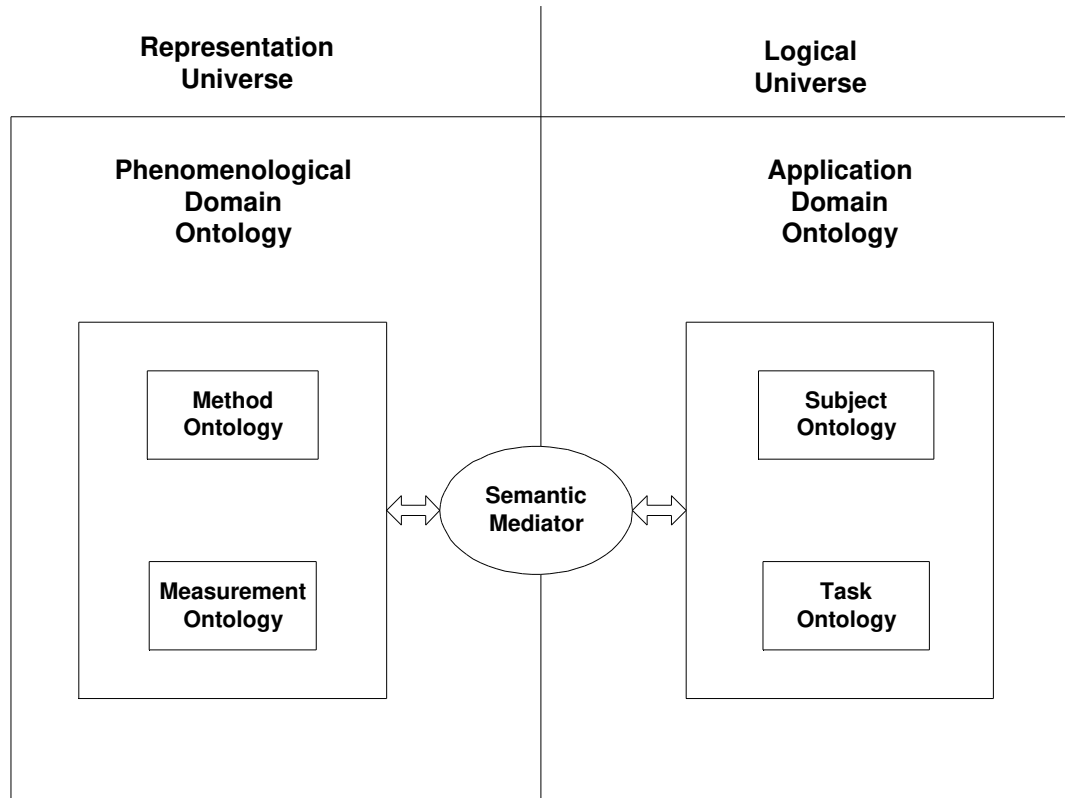
Figure 3-2 Phenomenological and application ontologies

One of the objectives of separating geo-ontologies in PDO and ADO is to emphasize the detection of *spatio-temporal configurations* of geographic phenomena. In a single time instance, the set of matchings of a concept from the application domain ontology to an instance of a concept on the phenomenological ontology is called a *spatial configuration*. Given a temporal sequence of geographic phenomena, the set of spatial configurations is called a spatio-temporal configuration. This idea is consistent with the identity-based modeling of change (Hornsby 1999), where object identity is proposed as a central notion for modeling spatial-temporal change. The

framework allows an object, identified as part of the user ontology, to be related to different descriptions in the PDO, because of changes in the object during a time series. Consider for example mapping urban sprawl for a city by analyzing a 20-year time series of LANDSAT images. The geometries that describe the evolution of the urban boundaries of the city change annually, yet the identity of the object remains the same.

Another objective is to be able to reuse elements of the same ontoloy in different applications. With this separation we make clear what are the specific methods and what are the more general ones. The specific methods can be reused for similar phenomena, while the general ones have a broader use. A simple example is the case of detecting or extracting line segments from a series of images. *Line segment* is a concept that is part of the structural ontology of the image. It has clearly defined geometric properties. These lines can take different roles in domain ontologies of different user communities. Another example is that all the methods for spatial analysis over polygons available on the PDO side can be reused for every application on the ADO side.

Each geographic object is unique as a concept in the logical universe and above. Although we choose different conceptualizations to represent it–*objects* and *fields*–its nature does not change. For instance, a reservoir is a reservoir, either represented by an aerial photograph, a vector representation, or a digital terrain model. Figure 3-3 shows a reservoir represented in three different ways.
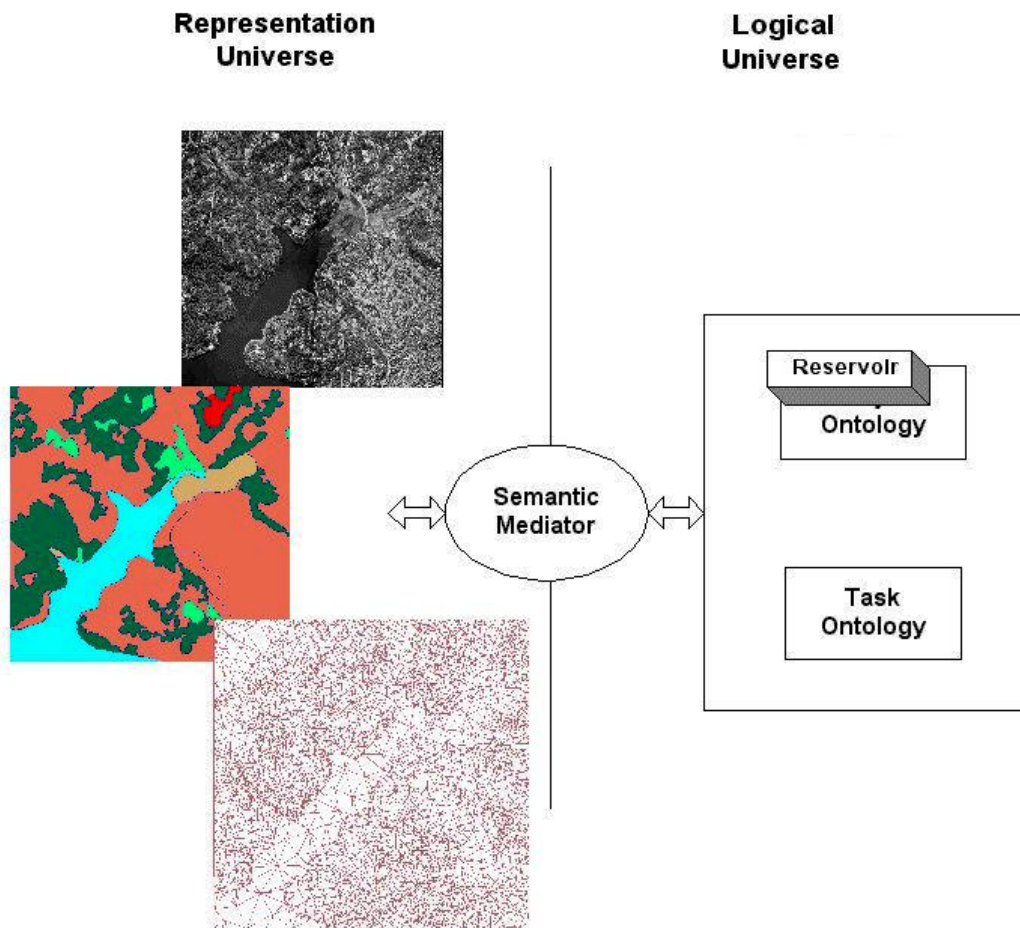
Figure 3-3 Three different representations of reservoir.

The representations are located in the representation universe, while the concept and its formal description are located in the logical universe. The concept reservoir is described only once in a high-level ontology, for instance a natural-geography ontology but it can be linked to more than one element in the PDO, (i.e., one for each of the different representations mentioned above).

3.2.1 Phenomenological Domain Ontology

An ontology of the geographic phenomena at the representation level has special characteristics, such as dependence on measurement, intrinsic properties, and reuse of algorithmic knowledge.

38

There is a strong dependency on the measurement process for both objects and fields. Objects recorded as collection of points can carry several differences. Points collected with a Global Positioning System (GPS) receiver need post-collection processing that can enhance the precision or dismiss some of the points (Figure 3-4).

<div align="center">(a)   (b)</div>
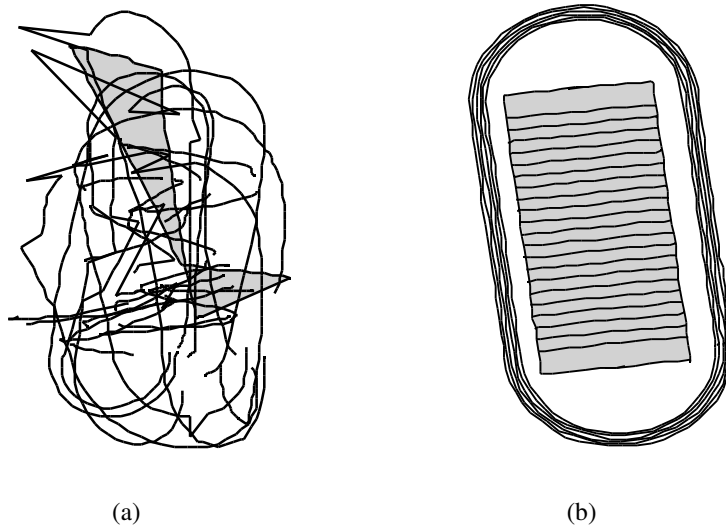
Figure 3-4 Lines and tracks of an athletic field collected with a GPS receiver (a) before and (b) after processing.

Data acquired by different sensors will also need different processing and give different results of the same phenomenon. Figure 3-5 shows an area in the Brazilian Amazon forest obtained by LANDSAT TM (optical) and RADARSAT L-band (radar). In the LANDSAT image it is possible to claim the existence of world objects (e.g., forest, as well as deforested and regrowth areas), whereas in the radar image is it more appropriate to consider the existence of land cover patterns, which result in different textures in the image. In fact, a large number of radar image classification algorithms are texture-based relying on the detection of statistical and structural texture measures.
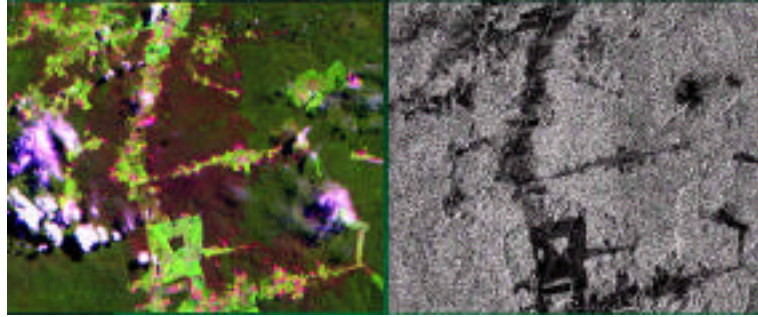
Figure 3-5 Two images of the same area captured differently.

Regarding intrinsic properties we argue that remotely sensed imagery cannot be reduced to the case of a single-date, single-band raster geometry, since most real-world uses of remotely sensed data rely on their temporal and multispectral nature. Image ontologies should consider their intrinsic properties: temporal cycle, multispectral capability, and spectral resolution.

Reusing algorithmic knowledge is important, because there is a significant amount of knowledge for different applications in the form of image processing algorithms, such as principal components, maximum-likelihood classifier, and texture measures. Applications that use the same kind of data will be able to reuse previously acquired knowledge.

The phenomenological domain ontology is measurement-dependent and has two distinct, but interrelated components:

- A measurement ontology describes the physical process of recording a geographic phenomenon. This recording process generates fields and objects. Regarding fields, there is the example of images where we are interested in expressing knowledge about the relation between energy reflected by the Earth's surface and the measurements obtained by the sensor. Typical concepts here include spectral response, backscatter, and Lambertian target. For objects, we are interested in the techniques for collecting the points, lines, and polygons that represent them. Procedures regarding precision and accuracy are also described here.

- A method ontology consists of a set of algorithms and data structures, which represent reusable knowledge to operate on the measured phenomenon. Sometimes they are in the form of processing techniques that can be used to transform the measured phenomenon from the representation level (e.g., by filtering or enhancement for images and geocoding for lines or polygons) to the logical level, or to perform feature extraction, segmentation, and classification in images leading also to the logical level. The operations regarding topology are also at this level, for instance, point-in-polygon operations, and the 9-intersection model.

The algorithms that are part of the method ontology perform transformations from the representation level to the logical level through a process called structural identification. When applied to an image (or a set of images), this process results in a set of structures strongly related to the measurement device properties and its interaction with the physical landscape. These structures may be geometric (e.g., regions extracted by a segmentation procedure) or functional (e.g., NDVI estimates obtained from NOAA/AVHRR series of images). When applied to objects, the result is the identification of the object and the association of this object to a meaning. For instance, what was just a polygon then becomes a `lake`, if it is associated with the class `lake` in an ontology present at the logical level.

3.2.2   Application Domain Ontology

Although the phenomenological ontologies are observer-independent, the application domain ontologies are not, because the domain scientists do their work using concepts from their knowledge domains. Within the application domain ontology we distinguish between two kinds of ontology: a subject ontology, which describes the vocabulary related to a generic domain (e.g., geology or ecology), and a task ontology, which are specializations of a domain ontology, describing a task or activity within a domain, such as water pollution assessment for ecological studies.

The concepts in the ADO are able to deal with the phenomena independent of representation. For instance, take the example of a study of homicide rates in the city of São Paulo, Brazil (Figure 3-6). The study itself and its ontology are independent of

how the phenomenon is spatially represented. In this case the available data was registered by regions. The researcher wanted to work with a field-like distribution, because the study made more sense with a smooth distribution. Therefore, the researcher used geostatistics techniques to obtain a new map with the same data but represented differently.
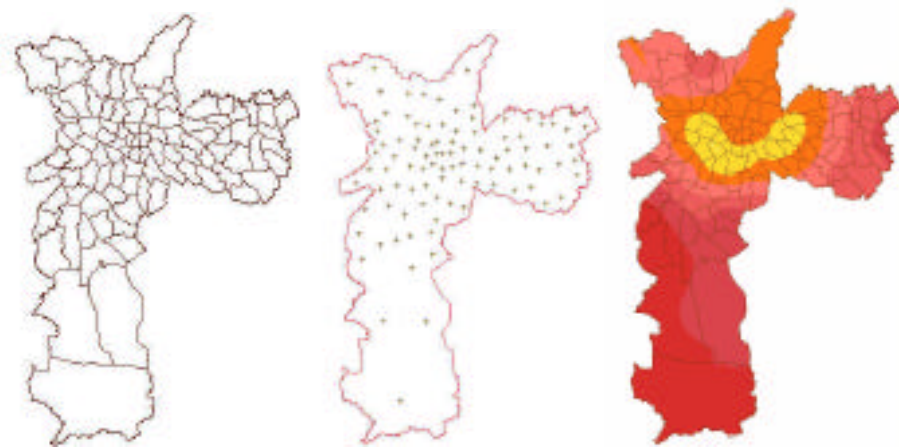
Figure 3-6 Three representations of the same phenomenon.

3.2.3 Semantic Mediators

The domain and task ontologies of the domain scientist include two different types of spatial entities: classes of identifiable objects that are modeled as objects and classes of spatially continuous phenomena that are modeled as fields. The relation between the phenomenological domain ontology and the application domain ontology is achieved by means of a semantic mediator, which performs two basic functions, i.e., *selection* and *identification*.

Selection is the operation of choosing the right methods that perform the identification. Image processing and pattern recognition algorithms described in the method ontology are needed to extract the desired structures from the image or to transform the physical values (i.e., pixel) into structures described in the structural ontology to obtain the desired information. For objects, this process is usually known as geocoding. For instance, to associate a set of centerlines to their correct codes there

is a series of methods, such as associating by individual addresses, zipcodes or street names.

Identification is the process of transforming generic entities present at the representation level, such as generic objects or images, into objects or image regions that have an identity. This process is a mapping from concepts on the subject ontology onto structures extracted from the image set. For example, a subject ontology may contain a concept of a road. Using the semantic mediator, we may try to identify linear structures in the representation level that correspond to roads at the application domain ontology (logical level). It is necessary to find the appropriate entities in the ontologies available at the logical level and make the association between them and the specific objects. Another example of identification is linking the gray area in Figure 3-4 to the football field in the Alumni Stadium and the lines to the Beckett Track.

The use of a semantic mediator allows different application domain ontologies to be related to a single phenomenological ontology, a perspective that reflects the fact that the same representation of a geographic phenomenon can be used in many knowledge domains. For example, the same set of images can be used for land-use and land-cover mapping or for geological studies.

There are many different ways one might create a semantic mediator. In this thesis, we consider the following constructive approach: an external observer builds the semantic mediator by forming a correspondence between concepts in the application domain ontology and measurements in the phenomenological domain ontology.

Consider the example of mapping deforestation of a tropical forest (Figure 3-7). In this image a segmentation algorithm has extracted regions from the pixel values (Shimabukuro *et al.* 1998). Two distinctly different types of deforestation can be observed: regular square-like patterns resulting from large cattle ranches and irregular patterns, resembling fish bones, which result from colonization projects. In this case, the subject ontology may distinguish generic types of concepts, such as forest, non-forest vegetation, and deforested areas. This latter concept could be specialized into

cattle ranches and small farms. At the phenomenological domain ontology level, we may distinguish such concepts as region and its specializations fishbone region and regular region. In the image, each region will be described by a set of statistical and morphological properties.

Figure 3-7 Deforestation mapping with a LANDSAT image (source: INPE).

A mapping between a concept in the subject ontology (e.g., small farm) and an instance of a concept on the measurement ontology (e.g., one instance of fishbone region) defines a matching. The set of all matchings between instances of fishbone regions to small farm defines, for this specific image, is a spatial configuration. When this set of matchings is performed in a time-series of images containing deforested regions, the set of spatial configurations of matchings of fishbone regions to small farms is a spatio-temporal pattern.

3.3 Bi-Directional Integration

One of the main objectives of this thesis is to integrate geographic information from different sources. The diverse geospatial information communities have different views of the world. These views can be formalized in different ontologies. Therefore, it is necessary to accommodate multiple ontologies, which in our model lie both inside the logical universe and inside the representation universe.

We introduce here two different ways to integrate ontologies. The first is the integration inside one subject and is called *vertical integration*. The other kind of integration is called *horizontal integration*, and involves integrating ontologies of different subjects (Figure 3-8).
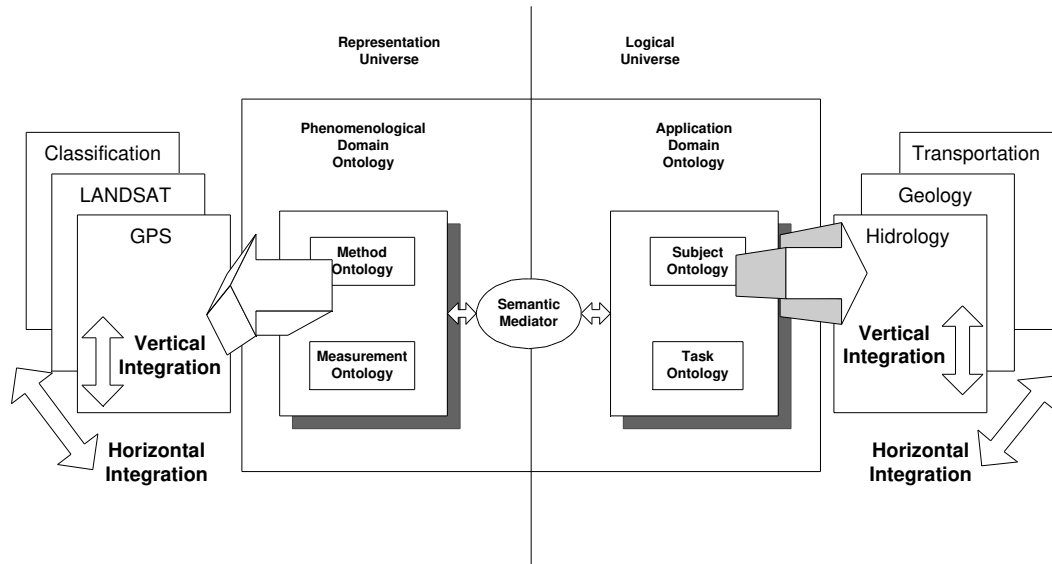
Figure 3-8 Horizontal and vertical integration.

When a new ontology is specified it is necessary to have a set of operations that allow the reuse of previous ontologies or parts of them. In an ODGIS environment three operations are available: *inheritance*, *inclusion*, and *roles*. Inheritance is used for vertical integration and roles are used for horizontal integration. Inclusion can be used for both integrations.

Classes in ODGIS are defined hierarchically, taking advantage of *inheritance*. It is possible to define more general classes, containing the structure of a generic type of object, and then specialize these classes by creating subclasses. The subclasses inherit all properties of the parent class and add some more of their own. For instance, within a local government there may exist different views and uses for land parcels. A standardization committee can specify a land parcel definition with general characteristics. Each department that has a different view of a land parcel can specify

its own land parcel class, inheriting the main characteristics from the general definition of land parcel and including the specifics of the department. In this way, a land parcel is defined for the whole city and derive two different specializations, one for tax assessment and the other for building permits.

We use *roles* to get around problems with multiple inheritance. In multiple inheritance for instance, a geographic feature can be at the same time a lake and a tourist attraction. In ODGIS we represent this entity as a lake that plays a role of a tourist attraction. Maybe later the lake can be considered as an environmentally protected area, that is, another role played by the entity lake. In ODGIS an entity can have many roles.

Inclusion is an operation in which an entity of one ontology is used to specify any part of an entity in a new ontology. For instance, an ontology that deals with representations of spatial objects will include many parts from a geometry ontology.

The integration operations are used in different stages of the ontology specification process. This separation happens because the levels of detail are different at the many stages of ontology specification. We suggest the use of inheritance in the high-level ontology integration and inheritance and roles at the low-level integration. Inclusion is used in every level of integration.

The multi-level ontology approach generates a very flexible model. In order to exploit this flexibility, we need a specific model for navigation among the diverse entities. We choose to develop the navigation model in the implementation universe. Since the classes extracted from the ontologies are in this level, the navigation model is based on change of classes.

3.4   Summary

This chapter introduced a multiple-ontology approach to geographic information integration. The different kinds of ontology, phenomenological domain ontology and

application domain ontology, were introduced. The operations inheritance, inclusion, and roles that are used for navigation inside ontologies were presented.

Chapter 4

A Methodology for Creating an ODGIS

The use of ontologies translated into active information system components leads to Ontology-Driven Information Systems (ODISs) (Guarino 1998) and, in the specific case of GIS, it leads to Ontology-Driven Geographic Information Systems (ODGISs) (Fonseca and Egenhofer 1999). ODGISs are built using software components derived from various ontologies. These software components are classes that can be used to develop new applications. Being ontology-derived, these classes embed knowledge extracted from ontologies.

The ODGIS framework is presented in the next two sections focusing on the aspects of knowledge generation and knowledge use (Figure 4-1). First we show how the ontologies are specified by the geospatial communities. Then we present how the knowledge generated in the first phase of the system can be used to develop GIS applications. This chapter also presents the mechanism for changes of classes. This mechanism allows an instance of a class to be generalized or specialized thus enabling information integration at different levels. The different levels of information granularity and their relation to different levels of ontologies are discussed here. The navigation introduced here shortens the gap between generic and specialized ontologies, enabling the sharing of software components and information.
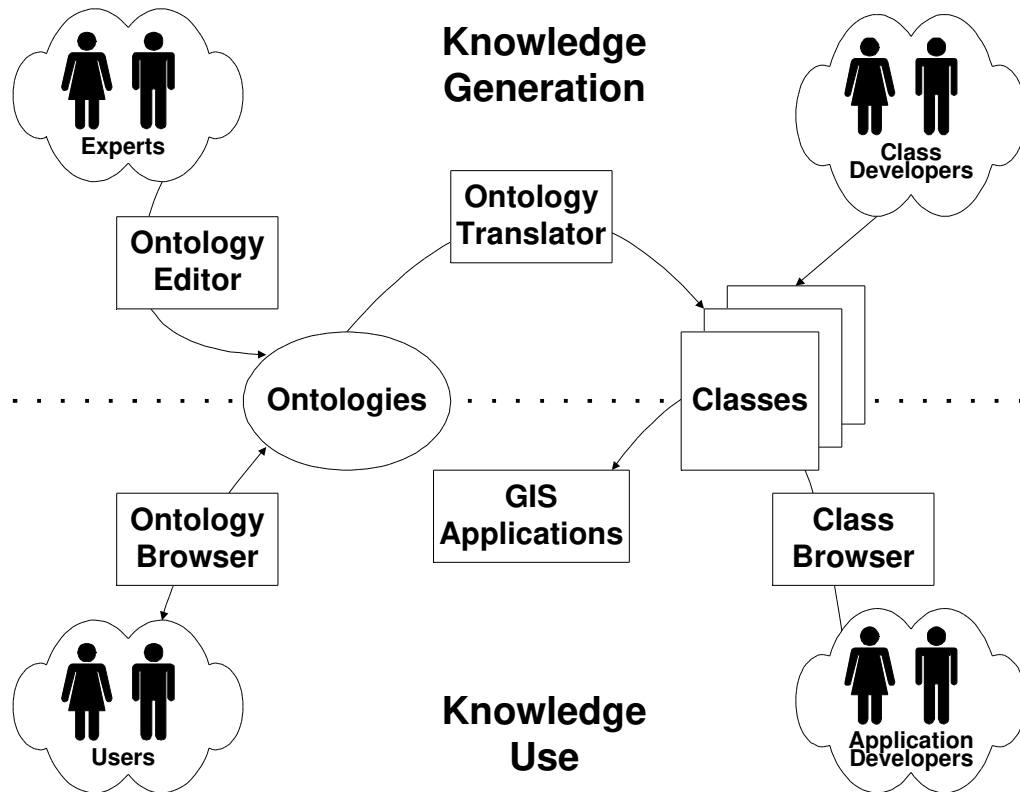
Figure 4-1 ODGIS framework.

4.1   Knowledge Generation

Ontology-driven geographic information systems are supported by two basic notions: (1) making the ontologies explicit before the systems are developed and (2) the hierarchical organization of communities. Explicit ontologies contribute to better information systems because every information system is based on an implicit ontology. The act of making the ontology explicit avoids conflicts between the ontological concepts and the implementation. Furthermore, top-level ontologies can be used as the foundation for the integration of systems, because they represent a common vocabulary shared by many communities.

Sometimes there is a misuse of the terms ontologies and database schemas. We are discussing here ontologies and not database schemas. Ontologies are semantically riches than database schemas and thus closer to the user's cognitive model. Our

approach is based on a group of people reaching an agreement about the basic geographic entities of their world. It does not matter whether the entities are stored in a database. A database schema represents what is stored in the database. An ontology represents concepts in the world. Although ontologies and database schemas can be related, ontologies are richer than database schemas in their semantics. The ontologies we deal with are created from the world of geographic phenomena. The information that exists in the databases has to be adapted to fill in the classes of the ontologies. For instance, the concept of lake can be represented differently in diverse databases, but the concept is only one, at least from one community's point of view. This point of view is expressed in the ontology that this community has specified. In the ODGIS architecture, diverse mediators have to act to gather the main aspects of lake from diverse sources of information and assemble the instance of a lake according to the ontology.

The world is divided into different groups of people each one with a different view of the world. We use the term *geospatial information communities* (GIC) to name these groups. Each GIC is a group of users that shares an ontology of real world phenomena. It is a basic assumption of this thesis that ontologies of diverse user communities can be explicitly specified. These groups generate ontologies of different levels of detail, the broader the group the more general the ontology. For instance, in a city, the mayor and his/her immediate staff view the city at a given abstract level. The department of transportation has a different and maybe more detailed view of the city. Inside the department of transportation, the section in charge of the subway system will have an even more specialized view of the city. We consider shared ontologies as the high-level language that holds those communities together. For instance, the department of transportation of a big city has a software specialized in transportation modeling beyond the regular GIS package, and therefore, they use more than one data model. But the conceptualization of the traffic network of the city may be the same and so only one ontology can hold this conceptualization. A GIC may commit to several ontologies. The users have the means to share information through the use of common classes derived from ontologies. The level of detail of the information is related to the abstraction level of the ontology. We use shared ontologies in a flexible

50

way, as users can derive more specific ontologies from shared ontologies creating new ones that apply directly to their work. These more specific ontologies have been called *applied ontologies* (Smith 1998) and *application ontologies* (Guarino 1998). Since we propose a flexible approach that integrates multiple ontologies, the communities of users are not constrained by a single ontology, but instead they can use the shared ontologies as a link to other user communities. The deeper we go into the user ontology (i.e., in the ontology hierarchy), the less information the users will be able to share with other user communities.

In ODGIS it is necessary that GICs assemble and specify ontologies at different levels. The first ontology specified inside a community is a top-level ontology. The assumption here is that this ontology exists and that it can be specified. The question of whether this *one ontology* exists or not is a matter very discussed and on which no consensus exists. We argue that it exists inside each community, although it can be sometimes too generic. People inside each community communicate, and therefore they agree on the most basic concepts. The top-level ontology describes these basic concepts. Specific ontologies can be created after the top-level ontology is specified. Medium-level ontologies are created using entities and concepts specified in high level ontologies. These concepts are specified here in more detail and new combinations of entities can appear.

For instance, consider a concept such as *lake*. It is a basic assumption of this thesis that a consensus can be reached by a GIC about which are the basic properties of a lake. Mark (1993) agrees that a generic definition of a class can be specified by its most common properties and thus avoid a rigid definition of exactly what a lake is. More specific definitions can be made at lower levels. This idea is applied in our multi-level ontology structure. We also share the belief of Smith (1998) that these different high-level concepts will converge on each other leading to common ontologies. The mechanisms introduced by this thesis enables the sharing of the common points of these theories.

A lake can be seen differently by different GICs. For a water department a lake can be a source of pure water. For an environmental scientist it is a wildlife habitat. For a tourism department it is a recreation point. The Wordnet ontology (Miller 1995) and the ontology extracted from SDTS (USGS 1998) were combined (Rodríguez 2000) and the result is used in the example. In this combined ontology a lake is "a body of (usually fresh) water surrounded by land." SDTS can be considered as a high-level ontology. Other concepts of lake can be derived from this high-level ontology. This is done using inheritance. The new concepts of lake will have all the basic properties defined in the Wordnet-SDTS ontology plus the add-ons that the GIC think are relevant to their concept of lake. The same happens with the other GICs. If they all are derived from the WordNet-SDTS lake they will be able to share complete information at this level only, although at a lower level they will share partial information.

There are two options to build the ontologies. First, we can consider that small GICs can assemble with other GICs with the same interests and try to build from their existing ontologies a high-level ontology that encompasses their lower level ontologies. The second option is that these GICs assemble before specifying their own ontologies in order to specify a high-level ontology for these group of communities. The most important thing here is that the architecture of an ODGIS allows reusing and integration of ontologies based in the reuse of classes through the use of inheritance and roles. The same rationale applied inside one community can be expanded to high level communities or to subgroups inside a community.

4.2   Knowledge Use

The result from the work of the GICs with the ontology editor is a set of ontologies. Once the ontologies are specified they can be translated to classes. The translation is available as a function of the ontology editor. The ontologies can be browsed by the end user, and they provide metadata about the available information. The set of classes contains data and operations that constitute the system's functionality. These classes

contain the knowledge available to be included in the new ontology-based information systems.

A basic schema of an ODGIS is (Figure 4-2):
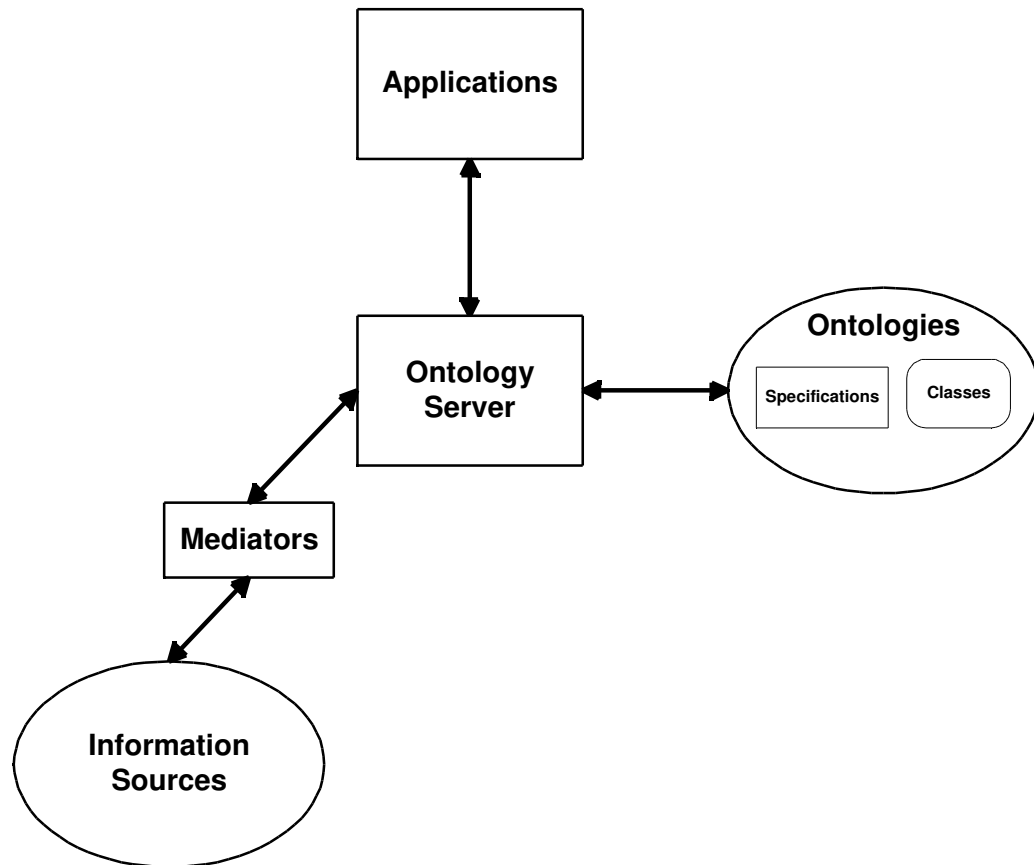
Figure 4-2 Basic components of an ODGIS.

- The *ontology server*: the ontology server has a central role in an ODGIS because it provides the connection among all the components. The server is also responsible for making the ontologies available to applications. The connection with information sources is done through mediators. Mediators look for geographic information and translate it into a format understandable by the end user. The mediators are pieces of software with

embedded knowledge. Experts build the mediators by putting their knowledge into them and keeping them up to date.

- The *ontologies*: they are represented by two kinds of structures, (i.e., the specifications and the classes). The specifications are made by the experts and stored according their distinguishing features. The structure provides information about the meaning of the available information. It can be used by the user to know which information is available and to match his/her conception of the world with other available conceptions stored by the ontology manager. The classes are the result of the translation of the ontologies.

- The *information sources*: the sources of geographic information in an ODGIS can be any kind of geographic database as long as they commit themselves to a mediator. The mediator has the function of extracting the pieces of information necessary to generate an instance of an entity belonging to an ontology. The mediator also has the function of bringing back new information in the case of an update.

- The *applications*: the main application of an ODGIS is information retrieval. The mediators provide instances of the entities available in the ontology server. The user can browse the information at different levels of detail depending on the ontology level used. Other kinds of applications can be developed, such as database update and different kind of geographic data processing, including statistical analysis and image processing.

4.3   Mechanisms for Changes of Classes

The knowledge-use phase of an ODGIS uses the products from the previous phase: a set of ontologies formally specified and a set of classes. The ontologies are available to be browsed by the end user and they provide metadata about the available information. A set of classes that contains data and operations constitutes the system's functionality. These classes are linked to geographic information sources through the

use of mediators. In this section we will discuss the operations of generalization and specification over the instances of classes from ontologies. The operations described here are applied over instances of classes, the real objects with data and operations.

There are two types of changes of classes in ODGIS. The first type occurs when an instance of a class immediately above or immediately below is generated from a given class. We call this transformation *vertical navigation*. The second type occurs when one of the roles played by the object is extracted from one instance. This way a new instance is generated producing a new object that belongs to the class of the role. We call this transformation *horizontal navigation* (Figure 4-3).
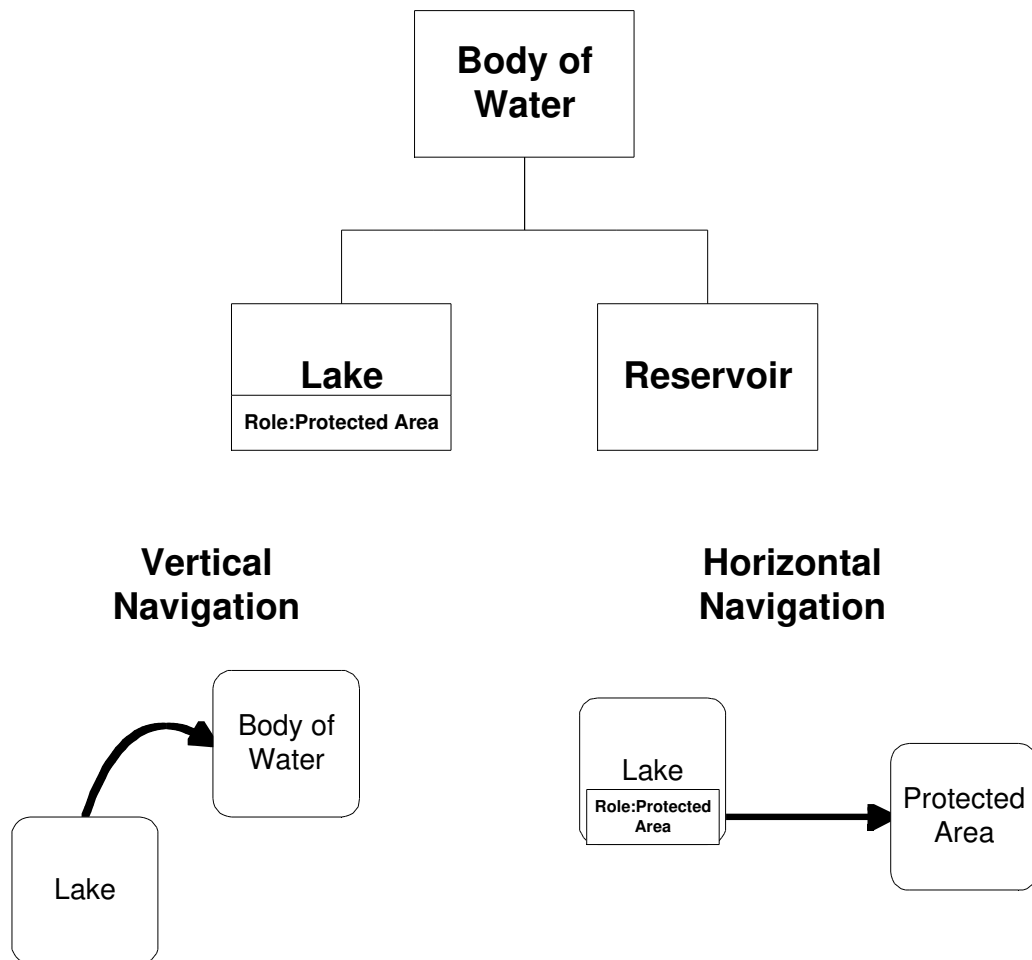
Figure 4-3 Vertical and horizontal navigation in an ontology of bodies of water.

Vertical navigation implies a change of level of detail, because it produces a new instance with more detail or one with less detail than the original instance. Horizontal navigation does not imply a change of the level of detail. The new class generated by horizontal navigation can be at any level in the hierarchy of classes.

4.3.1   Semantic Granularity in ODGIS

The abstraction of concepts and notions about real-world objects is an important part of the creation of information systems. In the abstraction process, certain characteristics of the objects are identified and coded in a database in such a way that the set of characteristics is representative of the much more complex real-world object. Depending on the user's interest, however, this set of characteristics can be defined to be more or less detailed.

Some authors consider *granularity* in a spatial database to be the same as resolution, thus implying that granularity is related to the level of distinction between elements of a phenomenon that is represented by the dataset (Stell and Worboys 1998). Hornsby (1999) points out the difference between resolution and granularity. Resolution refers to the amount of detail in a representation, while granularity refers to the cognitive aspects involved in selection of features. This kind of granularity is called *semantic granularity*. The notion of granularity applied to GIS leads to studies of the variation in the representation of geographic objects and phenomena across a wide range of scales. Certain phenomena are scale-dependent, (i.e., their representation varies across the scales). For instance, if an urban settlement is perceived at a small scale, the level of detail is usually small enough for an entire city with all its complex internal structure to be represented as a point or as a simple polygon on a map. If the same city is perceived at a larger scale it becomes necessary to represent its internal structure with more detail, for instance depicting blocks, squares, major streets, and buildings. Considering a geographic database where two representations of the same phenomenon have to coexist, Beard (1987) shows how it could be possible to maintain and update only the most detailed version of the objects and then to filter out unwanted detail to produce the less-detailed version. Here we

work with a higher level of abstraction dealing with information systems instead of databases. In an ODGIS, a concept can have more than one representation. For instance, the usual concepts about a river are independent of how it is represented, whether as a network for transportation or as an important element of the environment of a region. In an ontology, a river is defined first by its general meaning. More specialized ontologies deal with representation issues later.

In the ODGIS framework there are different levels of ontologies. Accordingly there are also different levels of information detail. Low-level ontologies correspond to very detailed information and high-level ontologies correspond to more general information. Thus, if a user is browsing high-level ontologies he or she should expect to find less detailed information. We propose that the creation of more detailed ontologies should be based on the high-level ontologies, such that each new ontology level incorporates the knowledge present in the higher level. These new ontologies are more detailed, because they refine general descriptions of the level from which they inherit.

We follow Hornsby's (1999) approach, considering that the level of semantic granularity is related to the level of ontology used. Ontologies can be used to specify how high-level abstractions relate to concepts of a lower level by establishing methods that help to implement rules and constraints. Guarino (1998) proposes a distinction between coarse and fine-grained ontologies, or on-line and off-line ontologies. A coarse ontology consists of a minimal number of axioms and it is intended to be shared by users that already agree on a conceptualization of the world. A fine-grained ontology needs a very expressive language and has a large number of axioms. Guarino concludes that coarse ontologies are more likely to be shareable and should be used on-line to support the system's functionality. On the other hand, fine-grained ontologies should be used off-line, because they are accessed eventually for reference purposes. Our solution allows the user to incrementally go from coarse to fine-grained ontologies on-line, thus eliminating the division between on-line and off-line ontologies.

4.3.2   The Mechanism for Changes of Granularity

There are two operations for changes in the level of detail: generalization and specialization. In generalization one class with a certain level of detail generates a new class with less detail. For instance, using Guarino's ontology (Figure 2-1), a `Geographical Region` can be generalized into a `Location`. Specialization is the operation in which a more general class is converted into a more specific class.

In ODGIS every class inherits from a basic class called `Object`. This specific class has two basic methods to be used in changes of granularity. One method is used to generalize new classes and it is called `Up()`, and the other is used to specialize classes and it is called `Create_From()`.

For example, if a user is dealing with instances of the class `lake` and of the class `reservoir`, the user can see and manipulate the instances of those objects as instances of `body of water`. This way the user is able to obtain better results in queries, retrieving more objects than if he had used only `lake` or only `reservoir`.

In specialization we can consider the same example but in a different order. The user has an instance of `lake` but he/she is interested in using some methods only available for the class `reservoir` or the user wants to combine in a detailed fashion the data available about the class `lake` with the data available about the class `reservoir`. The solution presented here allows the user to generalize first the instances of `lake` into `body of water`, and then from this new set of instances, specialize them into `reservoir`.

4.3.3   Generalization and Specialization

The generalization operation implies generating a new instance of a class with less detail and less knowledge than the original instance. To perform this operation it is necessary to have knowledge about which kind of data is going to be thrown out and which kind of data is going to kept or transformed. The best place to do this is inside

the instance that has all the data of the object that is going to be generalized. The operation that performs the generalization is called `Up()` and it implies changes not only to non-graphic data but also changes in representation formats. Generalizations of representation formats have been discussed elsewhere (Bertolotto and Egenhofer 1999; Davis and Laender 1999; Parent *et al.* 2000). What is presented here is the framework in which this kind of operation can happen. ODGIS is a framework that enables the integration of existing knowledge, either at the logical level or at the representation level.

The specialization operation implies generating a new instance of a class with more detail and, therefore, more knowledge embedded in it. In order to accomplish specialization we choose to place the method for specialization in the class that will receive the result of the operation. This choice was made because the know-how to perform this operation resides in the new class. Therefore, the class provides the methods and the rules for creating a new instance of itself from a more generic instance.

For example, if an instance of `reservoir` is going to be created, only the `reservoir` class knows all the details necessary to create an instance of itself. To create a class of `reservoir` from `lake` it is necessary that

- an instance of `lake` creates an instance of `body of water;`

- an empty instance of `reservoir` is created; and

- the instance of `reservoir` populates itself with data from the instance of `body of water.`

The result is an incomplete but working version of an instance of `reservoir`.

To make the instance of `reservoir` complete, the mediators have to look into the source of `reservoir` and then use similarity matching techniques (Rodríguez 2000) to try to match the new instance with available data. The result of this operation

is a more complete instance. From the point of view of `lake`, this new instance is richer, because it has all the information that it had before as `lake`, plus the information retrieved by the mediator from the source of `reservoir`.

4.3.4   Role Extraction

In an ODGIS, an object can play many roles. The object cannot change its own class without losing its identity, but it can play different roles depending on the context. In order to provide the user with the ability to work with these different roles we introduced the concept of *horizontal navigation*. This is the creation of a new instance of the class of the role played by an object. One of the roles played by the object is extracted, i.e., one new instance is available for the user.

This kind of operation is not an specialization or generalization, since a role can be seen as being at the same level of the originating classes instead of being at the level of their subtypes. For instance, a `lake` can play the role of a `link` is a transportation network. The ontology of bodies of water and the ontology of transportation can be at the same level.
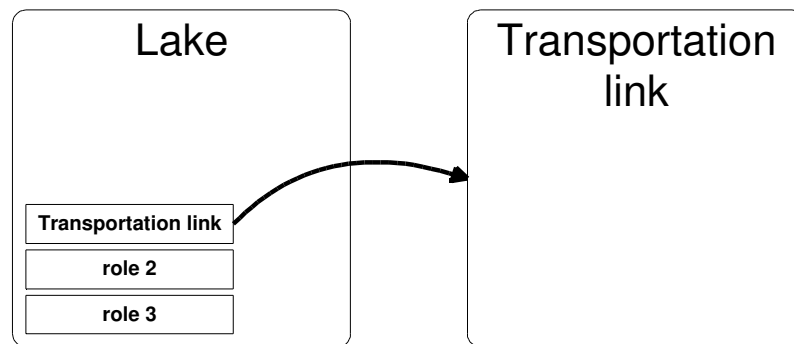
Figure 4-4 Role extraction.

The slots for roles are defined in the general class `object`. The rules and methods for generating an instance of a role should be provided in this class. The method for extracting a role is called `extract()`. For instance, the syntax to extract the role `link` from `lake` is: `new object link = lake.extract(link)`.

4.4   Summary

This chapter presented the ODGIS framework focusing on the aspects of knowledge generation and use. First it was shown how the ontologies are specified by the geospatial communities. Then it was presented how the knowledge generated in the first phase of the system can be used.

This chapter also presented the mechanism for changes of classes. This mechanism allows an instance of a class to be generalized or specialized thus enabling information integration at different levels. The different levels of information granularity and their relation to different levels of ontologies was discussed here. The navigation introduced here shortens the gap between generic and specialized ontologies, enabling the sharing of software components and information.

The next chapter of this thesis presents an assessment of alternatives for integrating ontologies.

Chapter 5

Ontology Integration

Geographic entities are complex objects that require sophisticated structures, such as ontologies, to abstract and represent them. A framework that intends to work with these kinds of objects should provide a solution for the problem of integrating ontologies. Furthermore, the development of a single theory, unifying all others is unlikely (Frank 1997) or at least will take some time (Smith 1998). To build an ontology of geographic objects it is necessary to integrate, for instance, a spatial ontology with a geometric ontology and a spatial reference system ontology. Section 5.1 presents the ODGIS view of information integration, and section 5.2 introduces the operations available for ontology integration. In the second part of the chapter, sections 5.3 through 5.6 present a measurement of the potential for information integration, discuss the experiment used to test the hypothesis, and show the results. The chapter's summary is in section 5.7.

5.1   Information Integration

The basic principle in this thesis is to allow for the integration of what is possible, instead of trying to integrate everything. It is our premise that once you achieve some kind of integration then further integration can occur incrementally. Some kinds of information will never be completely integrated since their natures are fundamentally different. For instance, a lake from the point of view of a parks and recreation department (`lake p&r`) has different functions and attributes than a lake from the point of view of a water department (`lake w`). The assumption in this thesis is that the lake is only one entity, but it is seen differently by different groups of people; therefore, a complete integration of all information available in these two (or more)

views is impossible, but the common characteristics can be shared. It is the integration of the common parts of concepts that we address here.

In order to integrate the common parts of shared concepts we propose a hierarchical representation of ontologies. The integration is always made at the first possible intersection going upward in the ontology tree. For instance, if both views of lake (`lake p&r` and `lake w`) are derived from the same entity `lake` in the WordNet-SDTS ontology, the possible integration is made at this level (Figure 5-1).
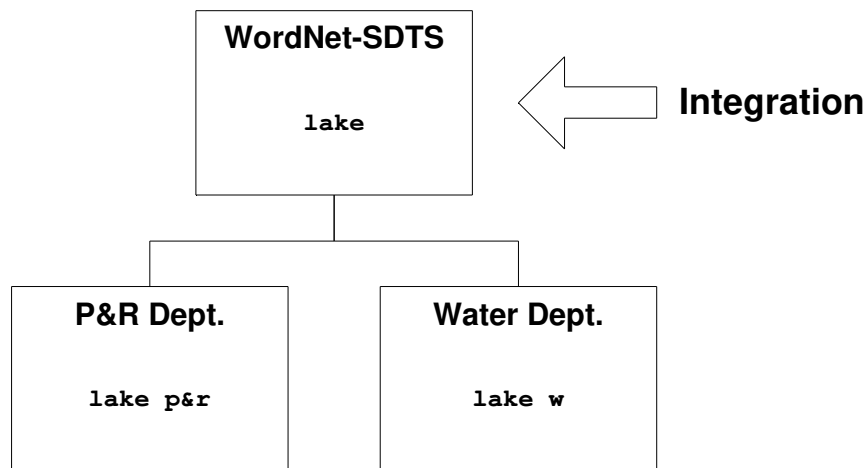
Figure 5-1 Integration of `lake.`

The integration includes all the methods and attributes of the class (i.e., the common methods and attributes of the class lake are all available for the user that is using the integrated information). In order for this to happen, it is necessary that the instances of `lake p&r` and `lake w` are converted to instances of the class `lake`.

In the similar way, roles can also be used to integrate information. A role in one class can be matched to another class or role. For instance, the role of a `wildlife habitat` that `lake` plays in the water-department ontology can be extracted and converted into an instance of `wildlife habitat` from the Environmental Protection Agency (EPA) ontology and then integrated with other instances of `wildlife habitat` coming from other sources of information.

The conversion of instances from one class to another is governed by a navigation method using the methods `Up()` and `Create_From()` inherited from the basic class called `Object`. These two methods provide the means to navigate through the whole ontology tree. Since each class in the ontology tree is derived from the basic class, each interface inherits the necessary navigation tools. So if the navigation method `Up()` is applied to `lake p&r`, the class returned is the next class in the upper hierarchy, the class `lake`.

5.2   Types of Ontology Integration

Ontologies can be integrated at different phases of a system life cycle. They can be considered for integration in (1) specification, (2) conceptualization, (3) formalization, (4) implementation, and (5) maintenance. When the integration happens in any of the first three phases we call it *high-level ontology integration* , and we call it *low-level ontology integration* when it happens in any of the last two phases.

The integration of ontologies in ODGIS is accomplished through the derivation of existing ontologies or through the insertion of existing ontology references into new ontologies. For instance, in an ontology for a parks and recreation department, the specification of `lake` can be inherited from `lake` in the Environmental Protection Agency (EPA) ontology. Different functions and attributes of a lake can come from different ontologies by use of the inclusion operation. The attribute location can also come from a different ontology. High-level ontology integration is done at the ontologies at the highest levels because these ontologies are stable, well developed, and are subject to few updates.

For instance, `lake` inheriting from `body of water`, `geographical region` inheriting from `region`, and `lake` having `beach` as one of its parts are considered high-level integration operations (Figure 5-2).
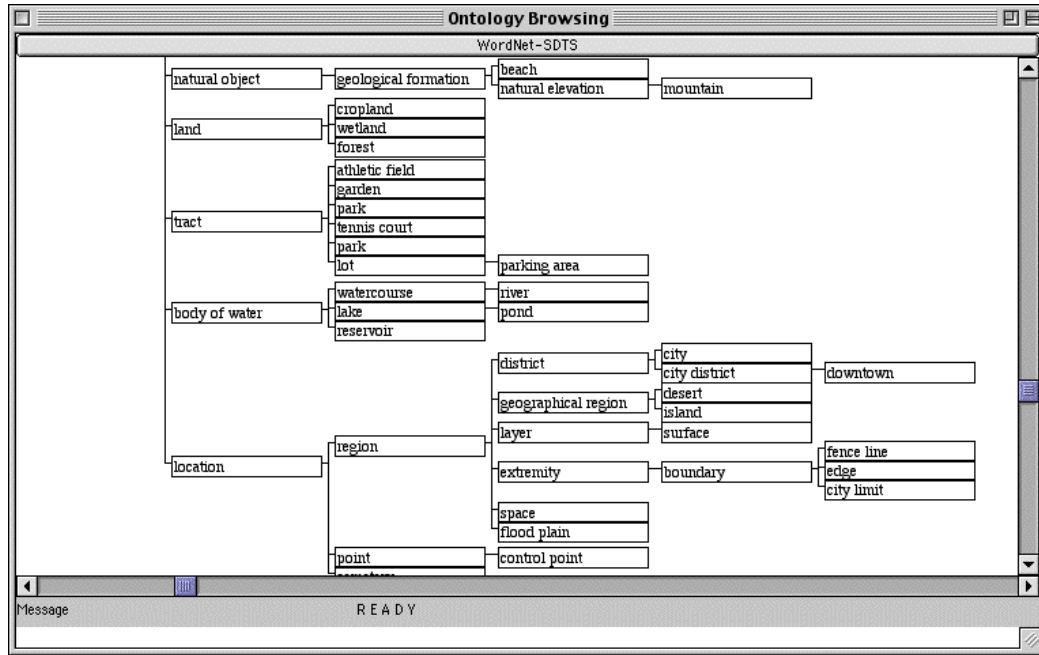
Figure 5-2 High-level integration.

At the level of an application ontology, more updates are expected. Application ontologies should be more flexible and be allowed to evolve with time. We propose that ontology integration at this level should be done through the integration of the classes. New classes are created through the use of inheritance. The new classes can play many roles that correspond to other classes in the ontologies. Since each role can come from a different ontology, the ontology integration is achieved through these classes. A `lake` that plays the role of `geographical region` or a `lake` that plays a role of `surface` are examples of low-level integration operations (Figure 5-3).
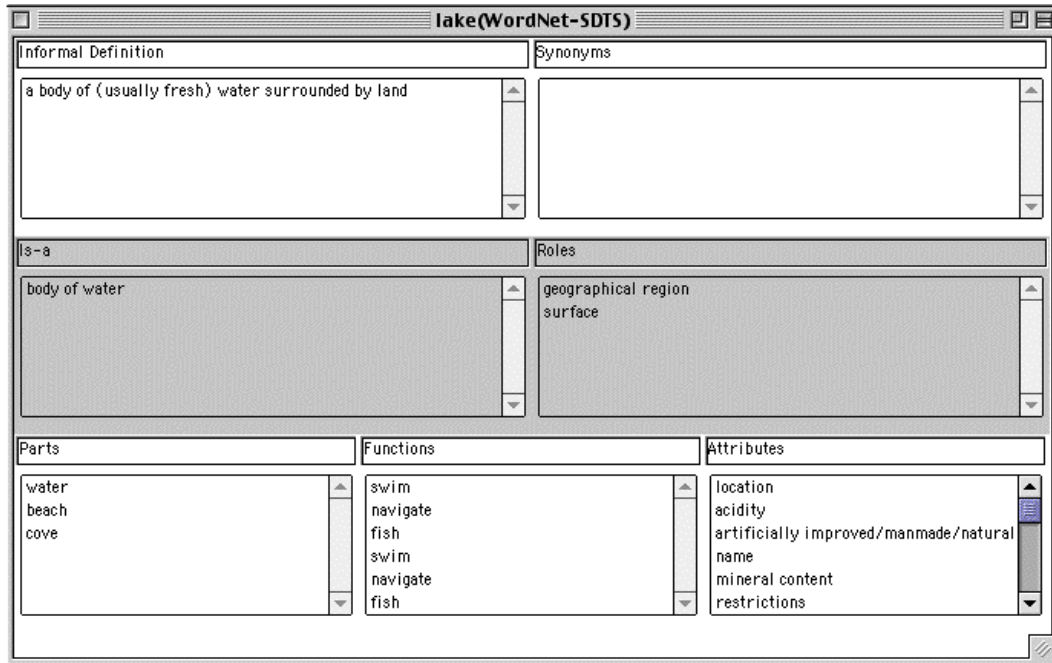
Figure 5-3 Low-level integration.

5.3 Measuring the Integration of Ontologies

In an ODGIS environment, when a user is trying to retrieve information from different sources it is necessary to combine the ontologies that represent the information. The multi-level approach used here allows for different levels of ontology integration. The entities in the ontologies are linked to sources of information. Therefore, if we measure the combination of ontologies we can evaluate the potential for information integration.

In order to measure the potential for information integration we took into consideration the kind of matching that happens at the entity level inside an ontology. When combining two ontologies the resulting potential for information integration is the sum of the potential for information integration of each match of an entity in one ontology to an entity in the other ontology, considering all the possible combinations. All possible matches are checked and the ones that can be accomplished are considered in the final result. Therefore, there is an evaluation for matches at the entity

level. The result of each match is accumulated giving a numerical measure for the potential for information integration when combining two ontologies.

One kind of integration that is possible is through the use of roles. One role in an entity can be matched to an entity in another ontology, or even to a role played by another entity. The possible matches are (Figure 5-4):

- one entity in the first ontology with one entity in the second ontology (E-E);

- one entity in the first ontology with one role in the second ontology (E-R);

- one role in the first ontology with one entity in the second ontology (R-E);

- one role in the first ontology with one role in the second ontology (R-R).
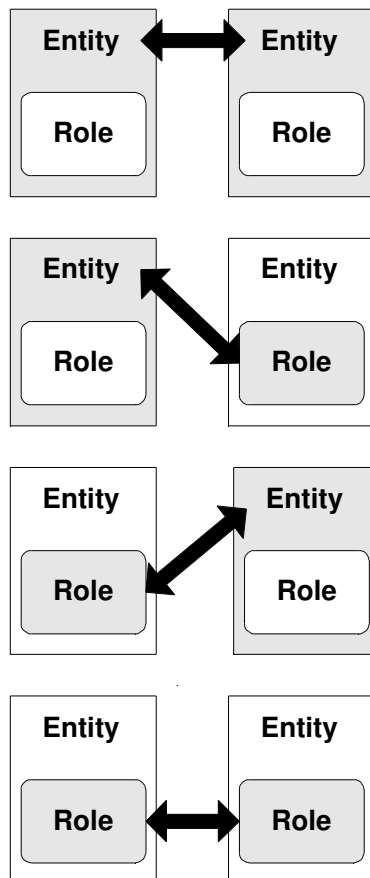
Figure 5-4 Types of integration using roles.

The second kind of integration is accomplished through the use of hierarchies. Since we use a hierarchical structure to represent ontologies, we can try to extend the possibilities of integrating information by including the parent of an entity in the search for a match. Considering the influence of hierarchies, the possible matches are (Figure 5-5):

- one entity in the first ontology with the parent of one entity in the second ontology (E-P_E);

- one role in the first ontology with the parent of one entity in the second ontology (R-P_E);
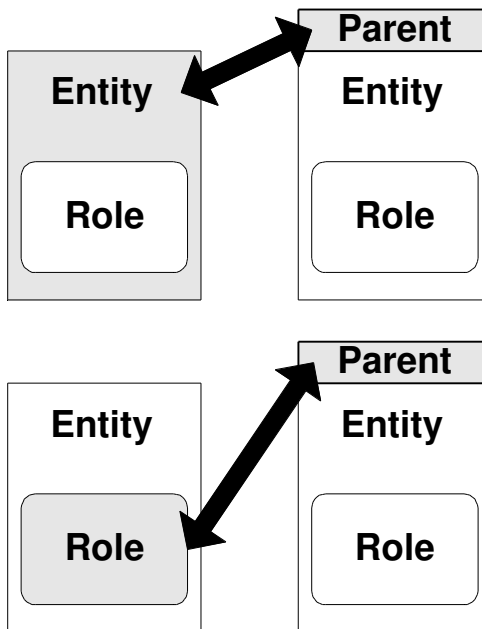
Figure 5-5 Types of integration using hierarchies and roles.

The final result reflects the sum of the amount integrated in all the matches of E-R, E-E, R-R, R-E, E-P_E, and R-P_E. A schema of what is computed is shown in Figure 5-6.
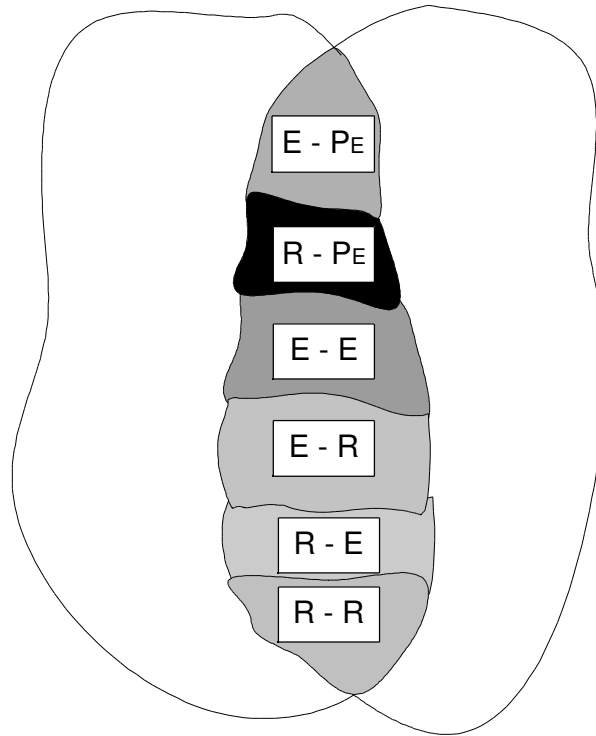
Figure 5-6 Possible matches between two ontologies: E-P$_E$ (Entity-Parent of
Entity), R-P$_E$ (Role-Parent of Entity), E-E (Entity-Entity), E-R
(Entity-Role), R-E (Role-Entity), and R-R (Role-Role).

5.4   A Method for Evaluating the Potential for Integration of Information

A domain can be represented as a set of ontologies (Equation 5.1).

$$\text{Domain D : } \{O_n\} \; n \geq 1 \tag{5.1}$$

An ontology can be represented as a set of entities that belong to a domain
(Equation 5.2).

$$\text{Ontology } O_n : \{E_i\} \; E \subseteq D, \; 0 \leq i \leq n, \text{ where n is the size of the ontology} \tag{5.2}$$

An entity can be represented as a set that includes an identifier and a set of roles
(Equation 5.3).

$$\text{Entity E: } \{id, R\} \tag{5.3}$$

The representation of an entity is much more complex than this. The term *id* includes everything that helps to identify uniquely an entity, such as the set composed of the definition, the parts, the functions, and the attributes.

A set of roles can be represented as a set of entities (Equation 5.4).

$$R_n: \{E_i\} \ E_i \subseteq D, 0 \leq i \leq n, \text{ where n is the size of the set.} \quad (5.4)$$

Calculating the potential of information that can be integrated between two ontologies is done by comparing each component of each ontology. The main factors in this operation are the entities, the roles, and the parent classes of each entity.

The potential for information integration in the matching of two ontologies is given by the sum of the potential for information integration in each pair of entities that can be formed through the combination of all entities in one ontology with all the entities in the other ontology. The general formula for measuring the potential for information integration when combining two ontologies O_n and O_m is given in Equation 5.5.

$$I = \sum compare_k (E_1, E_2) \ E_1 \subseteq O_n, E_2 \subseteq O_m, \text{ where k is n x m} \quad (5.5)$$

I is the number that gives the potential for information integration, *compare* is a function that matches the entity E_1 to the entity E_2, E_1 is an entity of the ontology O_n, E_2 is an entity of the ontology O_m.

In order to learn how the hierarchical organization of the ontologies and the use of roles influence the potential for information integration we develop four different types of evaluation. In the next sections we present the method to measure the potential for the integration of information when combining two ontologies (1) using roles alone, (2) using roles and hierarchies, (3) using hierarchies alone, and (4) without using roles or hierarchies.

5.4.1 Evaluation with Roles Alone

When comparing two ontologies with the goal of integrating them, we can extend the potential of information to be integrated by adding the roles that an entity plays to the arguments of the comparison. The potential for information integration in the match of two entities, each one from one different ontology, including the effect of the use of roles, is given in Equation 5.6.

$$I_R = E_E + R_E + R_R \qquad (5.6)$$

where

- I_R is the potential for information integration considering the effect of roles;

- E_E has a value 1 if the id of E_1 is equal to the id of E_2, 0 if the id of E_1 is not equal to the id of E_2;

- R_E is the number of roles of E_1 with the id equal to the id of E_2; and

- R_R is the number of roles of E_1 that are equal to any role played by E_2.

The total potential for information integration when combining O_n and O_m considering the effects of roles is given in Equation 5.7.

$$I = \sum (I_R)_k, \text{ where k is n x m.} \qquad (5.7)$$

For instance, the entity `lake`, playing a role of `protected area`, is compared with the entity `transportation link` that plays the role of `lake` in a second ontology (Figure 5-7).
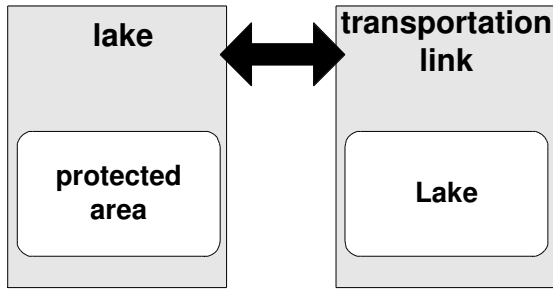
Figure 5-7 An entity vs. entity match.

The evaluation gives:

$E_E = 0$

$R_E = 1$

$R_R = 0$

$I_R = 0 + 1 + 0 = 1$

5.4.2 Evaluation with Roles and Hierarchies

If the ontologies are organized hierarchically we can increase the potential for information integration. Our framework allows the change of classes up and down in the hierarchy. Therefore, we can add to the basic evaluation the effects of the use of hierarchies in the representation of ontologies. By broadening the scope of comparison we can compare each entity and each role, not only with the matching entity and role of the other ontology, but also with the parent class of the matching entity. The potential for information integration in the match of two entities, each one from a different ontology, is given by the following formula, which includes the effects of the hierarchical organization:

$$I_{R+H} = I_R + I_H \tag{5.8}$$

The expanded formula is given in Equation 5.9.

$$I_{R+H} = (E_E + R_E + R_R) + (E_{E+H} + R_{E+H}) \tag{5.9}$$

From Equation 5.9 we have:

- I_{R+H} is the potential for information integration considering the effect of roles and of the hierarchy;

- E_E has a value 1 if the id of E_1 is equal to the id of E_2, 0 if the id of E_1 is not equal to the id of E_2;

- R_E is the number of roles of E_1 with the id equal to the id of E_2;

- R_R is the number of roles of E_1 that are equal to any role played by E_2;

- E_{E+H} has a value 1 if the id of E_1 is equal to the id of the parent of E_2, 0 if the id of E_1 is not equal to the id of the parent of E_2; and

- R_{E+H} is the number of roles of E_1 with the id equal to the id of the parent of E_2.

We consider here only the comparison of one entity (E_1) and its roles (R_{E1}) in the first ontology with each entity (E_2) of the second ontology and its parent (E_{2P}). If a match is not achieved with the immediate parent of E2, comparisons are made with the parent of the parent till the root of the ontology.

The total potential for information integration when combining O_n and O_m, considering the effects of roles and hierarchies, is given by Equation 5.10.

$$I = \sum (I_{R+H})_k, \text{ where k is n x m} \tag{5.10}$$

For instance, the entity `body of water` that plays the role of `protected area` is compared with the entity `reservoir` that plays the role of `protected area` in a second ontology. Furthermore, `reservoir` has `body of water` as its parent (Figure 5-8).
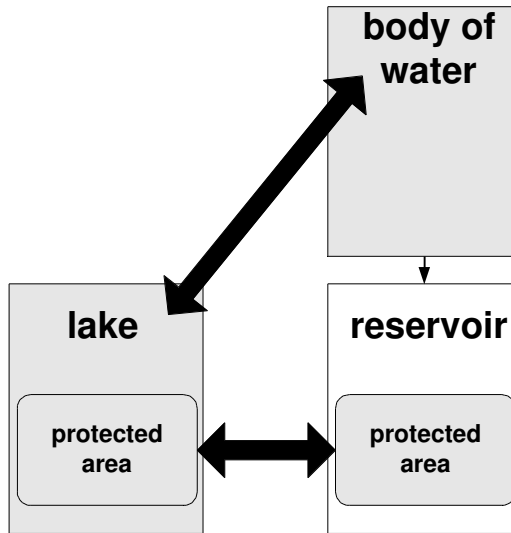
Figure 5-8 A mixed match.

The evaluation gives:

$E_E = 0$

$R_E = 0$

$R_R = 1$

$E_{E+H} = 1$

$R_{E+H} = 0$

$R_{R+H} = 0$

$I_{E+H} = (0 + 0 + 1) + (1 + 0 + 0) = 2$

5.4.3   Evaluation with Hierarchies Alone

Another way to evaluate the potential for information integration when combining two ontologies is without taking into account the roles played by the entities. Such an evaluation depends only on the hierarchical organization of the ontologies and is based solely on the comparison of two entities at a time, disregarding any roles if they exist.

The potential for information integration in the match of two entities, each one from one different ontology, is given by equation 5.11.

$$I_H = (E_E) + (E_{E+H}) \qquad\qquad (5.11)$$

where

- I_H is the potential for information integration considering the effect of the hierarchy alone;

- E_E has a value 1 if the id of E_1 is equal to the id of E_2, 0 if the id of E_1 is not equal to the id of E_2; and

- E_{E+H} has a value 1 if the id of E_1 is equal to the id of the parent of E_2, 0 if the id of E_1 is different from the parent of E_2. If E_E is equal to 1, E_{E+H} should be 0, because the integration has already occurred at the entity level without using parents.

The measure for total potential for information integration when combining O_n and O_m, considering the effects of hierarchies only, is given by Equation 5.12.

$$I = \sum (I_H)_k, \text{ where } k \text{ is } n \times m \qquad\qquad (5.12)$$

For instance, the entity `lake` is compared with the entity `reservoir` in a second ontology. Furthermore, `reservoir` has `body of water` as its parent (Figure 5-9).
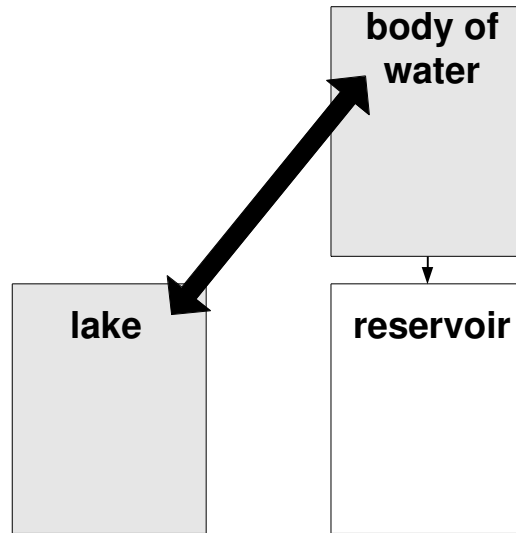
Figure 5-9 An entity vs. parent of entity match.

The evaluation gives:

$E_E = 0$

$E_{E+H} = 1$

$I_H = 0 + 1 = 1$

5.4.4 Evaluation without Roles and without Hierarchies

The simplest way to evaluate the integration of two ontologies in our setting is to disregard both the effects of the roles played by the entities and the hierarchical organization of the ontologies. Such an evaluation consists only of the comparison of two entities at a time, disregarding any roles if they exist and not making any comparisons with parent classes. The measure for the potential for information integration in the match of two entities, each one from one different ontology, is given by Equation 5.13.

$$I_{R-H} = (E_E) \tag{5.13}$$

where

- I_{R-H} is the potential for information integration without the effect of roles or the hierarchy; and

- E_E has a value 1 if the id of E_1 is equal to the id of E_2, 0 if the id of E_1 is not equal to the id of E_2.

The total potential for information integration when combining O_n and O_m is given by Equation 5.14.

$$I = \sum (I_{R-H})_k, \text{ where k is n x m} \tag{5.14}$$

For instance, the following entity `body of water` is compared with the entity `body of water` in a second ontology (Figure 5-10).
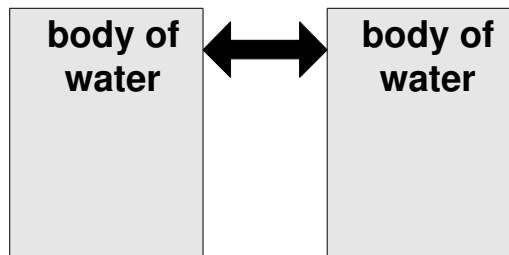
Figure 5-10 A simple match.

The evaluation gives:

$E_E = 1$

$I = 1$

5.5   The Simulation

The objective of the simulation was to model a geospatial information community. This community, a city for instance, has defined a basic ontology with a certain number of entities. Two departments of this city built their own ontologies based on this large set of entities. Now these two departments want to share information. So we need to integrate the two ontologies, one for each department. These two ontologies

may have no parts in common, or they may have some overlap. The objective of the experiment is to measure the intersection of the two ontologies, that is, the potential for information integration when combining two ontologies. The possible results of the combination of two ontologies are (a) no overlap at all, (b) small overlap, (c) large overlap, and (d) inclusion (Figure 5-11).

<div align="center">
a   b   c   d
</div>
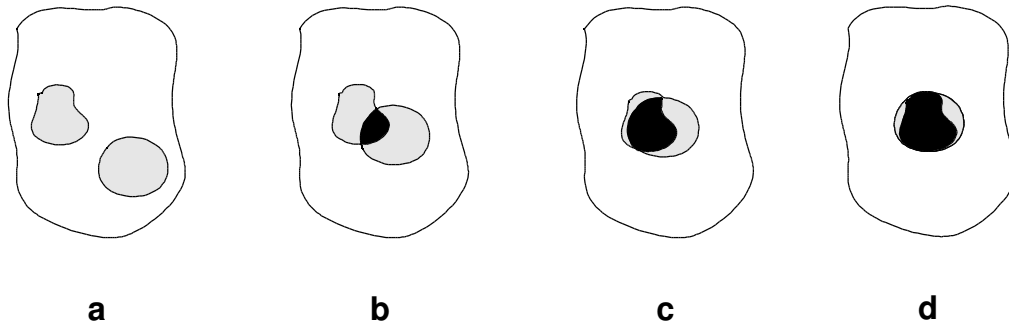
Figure 5-11 Possible results of the combination of two ontologies: (a) no overlap at all, (b) small overlap, (c) large overlap, and (d) inclusion.

In the experiment a large set of entities was randomly generated. Then two subsets of smaller ontologies were randomly extracted from the large set and compared with each other in order to measure the potential for information integration. The number of descendants of a given entity was randomly generated and varied from 0 to 5 descendants. The number of roles in a given entity was randomly generated and varied from 0 to 5 roles.

The results were normalized by dividing the measure of the amount of entities actually matched by the amount that could be matched. For instance, considering in the first ontology O_1 an entity E_1 with P_1 as a parent and playing two roles R_{11} and R_{12}. Considering the second ontology O_2 an entity E_2 with P_2 as a parent and playing two roles R_{21} and R_{22}. The maximum amount that can be integrated is given by the match of:

- E_1 with E_2, R_{21} or R_{22};

- plus the match of R_{11} with E_2, R_{21} or R_{22}; or

- the match of R_{12} with E_2, R_{21} or R_{22}.

In this example the largest amount of information that could be integrated is 3. If for instance E_1 is equal to R_{21} and R_{11} is equal to R_{22} then the evaluation of the match is 1 (E_1-R_{21}) plus 1 (R_{11}-R_{22}), summing up 2. The normalized result is 2/3.

5.5.1 The Small-Scale Experiment

This experiment simulates a community with an ontology of 1000 entities. There are two groups within this community that want to share information. This is a small community and, therefore, accommodates a small number of groups. The size of the ontologies of each group is 200 entities.

A set of 1000 entities was randomly generated for the community ontology. The number of descendants of an entity was randomly generated and varied from 0 to 5 descendants. The number of roles of each entity in this ontology was randomly generated, varying from 0 to 5. Two subsets of 200 entities were drawn from the larger set and compared for the evaluation of the potential for information integration.

We ran the experiment 100 times. The potential for information integration was recorded for four types of measurement: (1) using hierarchies and roles, (2) using roles alone, (3) using hierarchies alone, and (4) using neither hierarchies nor roles. A sample of the results is shown in Table 5-1. The results for the potential of information that could be matched for (1) and (2) are the same because the method for evaluation for both gives a maximum of 1 for the match of an entity with an entity or the match of an entity with a parent of an entity. The results for the potential of information that could be matched for (3) and (4) are the same because the method for evaluation uses the same rationale of (1) and (2) plus the effects of roles that are present in both.

Using neither hierarchies nor roles			Using hierarchies alone			Using roles alone			Using hierarchies and roles		
Actual matching	Could be matched	Normal-ized	Actual matching	Could be matched	Normal-ized	Actual matching	Could be matched	Normal-ized	Actual matching	Could be matched	Normal-ized
27	200	0.13	74	200	0.37	171	702	0.24	360	702	0.51
41	200	0.20	68	200	0.34	172	693	0.24	362	693	0.52
45	200	0.22	75	200	0.37	233	706	0.33	382	706	0.54

Table 5-1 Extract from the results of the small-scale experiment

For instance, for the measurement of the potential for information integration using roles alone we had a mean of 184.89 and a standard deviation of 24.48 for the non-normalized result. The normalized result had a mean of 0.36 and a standard deviation of 0.037. Table 5-2 shows the mean and standard deviation of the non-normalized and normalized measurements of the potential for information integration using no roles and no hierarchies, using hierarchies alone, using roles alone, and using roles and hierarchies.

	Using no roles and no hierarchies	Using hierarchies alone	Using roles alone	Using roles and hierarchies
mean of non-normalized results	36.64	72.1	184.89	368.52
standard deviation of the non-normalized results	5.62	7.47	24.48	25.52
mean of normalized results	0.18	0.36	0.26	0.53
standard deviation of the normalized results	0.028	0.037	0.034	0.030

Table 5-2 A summary of the results of the small-scale experiment.

The results of the potential for information integration are shown in Figure 5-12. The points in the graph represent the normalized results of each of the 100 times that the experiment was made. The Y axis represents the potential for information integration and the X axis represents the number of the experiment run.
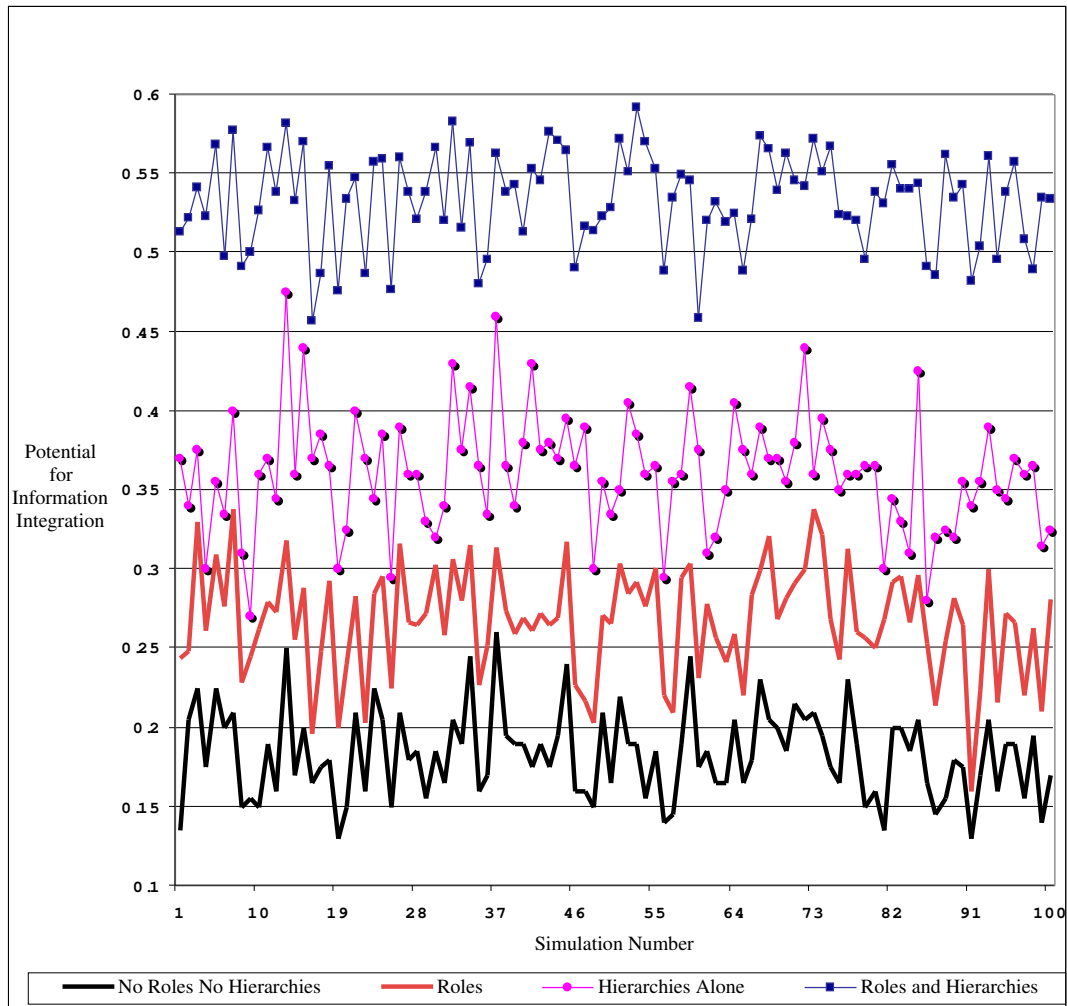
Figure 5-12 Graph results of the small-scale experiment.

The variation of each type of measurement is due to the randomness in the generation of the sets. The potential for information integration with hierarchies and roles was the greatest of all and had a mean of 0.53 and a standard deviation of 0.03. The potential for information integration with hierarchies alone had a mean of 0.36 and a standard deviation of 0.03. The potential for information integration with roles alone had a mean of 0.26 and a standard deviation of 0.03. The potential for information integration with no hierarchies and no roles had a mean of 0.18 with a standard deviation of 0.02.

5.5.2 The Large-Scale Experiment

The second experiment was made with the same rationale as the first one. The idea is simulating a ten times larger community. There are two groups within the community that want to share information. Since this is a large community, it accommodates more groups. The size of the ontologies of each group is 500 entities.

A set of 10,000 entities was randomly generated for the large community. The number of descendants of an entity was randomly generated and varied from 0 to 5. The number of roles of each entity in this ontology was randomly generated varying from 0 to 5. Two subsets of 500 entities were drawn from the larger set and compared for the evaluation of the potential for information integration.

The potential for information integration was recorded for four types of measurement: (1) using hierarchies and roles, (2) using only roles, (3) using only hierarchies, and (4) using neither hierarchies nor roles.

The results of the potential for information integration are shown in Figure 5-13. The points in the graph represent the results of 100 runs of the experiment. The potential for information integration with hierarchies and roles was the greatest of all and had a mean of 0.34 and a standard deviation of 0.019. The potential for information integration with hierarchies alone had a mean of 0.256 and a standard deviation of 0.017. The potential for information integration with roles alone had a mean of 0.0802 and a standard deviation of 0.012. The potential for information integration with no hierarchies and no roles had a mean of 0.049 and a standard deviation of 0.010.
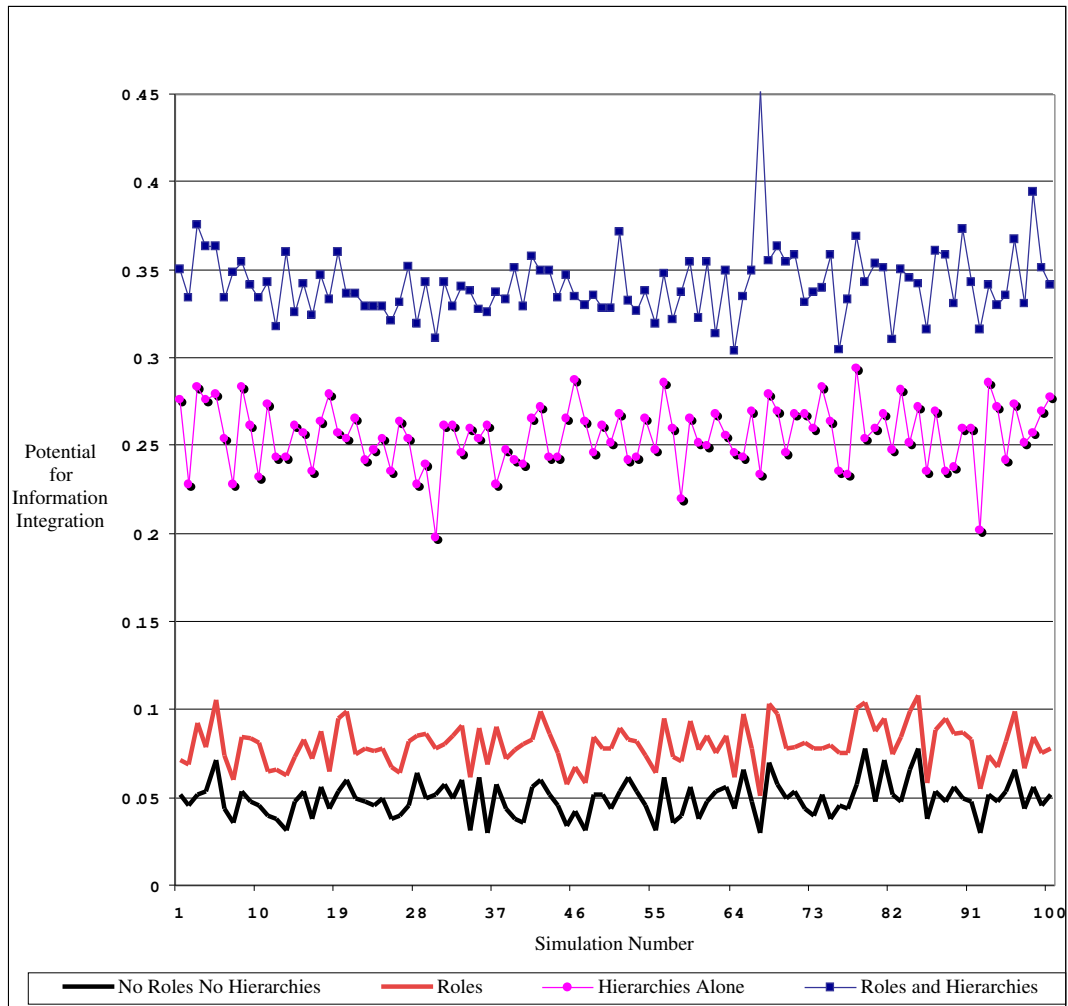
Figure 5-13 Potential for information integration in the large-scale experiment.

The non-normalized results for roles alone sometimes are greater than the results for hierarchies alone. But when these results are normalized the results for hierarchies alone are better. An extract of roles alone and hierarchies alone the first 20 times of the experiment is shown in Table 5-3.

	Roles alone	Hierarchies alone	Roles alone	Hierarchies alone
	non-normalized		normalized	
1st run	127	138	0.0718	0.276
2nd run	122	114	0.0689	0.228
3rd run	160	142	0.0930	0.284
4th run	139	138	0.0792	0.276
5th run	180	140	0.1063	0.28
6th run	123	127	0.0737	0.254
7th run	110	114	0.0601	0.228
8th run	143	142	0.0849	0.284
9th run	145	131	0.0837	0.262
10th run	145	116	0.0817	0.232
11th run	119	137	0.0656	0.274
12th run	114	122	0.0658	0.244
13th run	116	122	0.0628	0.244
14th run	127	131	0.0732	0.262
15th run	148	129	0.0831	0.258
16th run	126	118	0.0724	0.236
17th run	155	132	0.0876	0.264
18th run	110	140	0.0653	0.28
19th run	161	129	0.0956	0.258
20th run	169	127	0.0994	0.254
mean of 100 times	128.05	140.04	0.0802	0.256
standard deviation of 100 times	8.96	20.43	0.012	0.017

Table 5-3 An extract of the sample of the large-scale experiment.

5.6 Analysis of the Results

The results of both simulations were very similar. The potential for information integration is greater with roles and hierarchies than without them. The variation can be grouped into two types. A first group with the higher results is for ontologies that use roles together with hierarchies. The second group with lower results is the group in which the ontologies had hierarchies alone, roles alone, and no hierarchy and no roles.

5.6.1 The Effect of Using of Hierarchies

The use of a hierarchical structure to represent ontologies had a positive effect on the potential for information integration. Specifically, the potential for information integration with this model was better than the amount integrated with the model that used no hierarchy at all and better than the model that used roles alone.

5.6.2 The Effect of Using Roles

The use of roles in the representation of ontologies had a positive effect on the potential for information integration. Specifically, the potential for information integration using this model was better than the amount integrated with the model that used no roles and better than the model that used no roles and no hierarchies.

5.6.3 The Combined Effect

The combined effect of the use of roles and the hierarchical structure proved to be the best of all. In the small-scale experiment, the improvement over the model that used roles alone was 2 times more than the second best result. In the large-scale experiment it was almost 5 times better.

5.6.4 The Effect of Using no Roles and no Hierarchies

Of all the combinations, the potential for information integration with no roles and no hierarchies was the smallest. Using roles was 1.2 times better, using hierarchies was 1.6 times better, and the combined effect of roles and hierarchies was 3.7 times better than using no roles and no hierarchies.

5.6.5 Evaluation in Favor of Hypothesis

This experiment evaluated the influence of the number of roles and of the hierarchical structure for representing ontologies on the potential for information integration. We observed a strong influence of the use of a hierarchical structure in increasing the potential for information integration. The use of roles also improved the potential for information integration although to a much lesser extent than the use of hierarchies

alone did. The combined effect of roles and hierarchies had a more positive effect in the potential for information integration than the use of roles only or hierarchies only. All three combinations gave better results than the results using neither roles nor hierarchies. Therefore, we can say that the use of hierarchies and roles in the representation of ontologies increases the potential for information integration. This statement supports the hypothesis that a model that incorporates hierarchies and roles has a potential to integrate more information than models that do not incorporate these concepts.

5.7   Summary

This chapter addressed the issue of ontology integration inside the ODGIS framework. First the ODGIS view of information integration was presented. The high-level integration of ontologies was discussed, followed by a discussion oflow-level integration.

A method for evaluating the potential for information integration was introduced. The measurement captures the effects of the use of roles and of hierarchical structures in the representation of ontologies in the potential for information integration. Two experiments and their results, both of which supported the hypothesis, were described. The chapter also presented the conclusion that the use of roles and hierarchies have a positive influence in the potential for information integration. The use hierarchies alone had more influence in the potential for information integration than the use of roles alone.

The next chapter of this thesis presents guidelines for the implementation of the main components of an ODGIS.

Chapter 6

Guidelines for Implementation

In this chapter we analyze the options for implementation of the main components of an ODGIS. We are suggesting here specific tools for implementation. We know that these tools are not the only solutions but the evolution of ontology-driven information systems will lead to the use similar tools or to an evolution of these same tools.

An ontology-driven information system deals with instances of classes called objects. These objects are extracted from geographic databases and carry data and operations. One of the most suitable options for implementing interoperable objects or components (Betz 1994) that need to share both code and data across a heterogeneous network is the use the programming language Java (Clemens 1996; Lewandowsky 1998), because compiled Java code (bytecode) can be executed by Java interpreters available on most computers. Furthermore, the object-oriented structure of Java offers many features for the implementation of distributed objects. There are two options for implementing Java objects from ontologies. First, the objects can be generated from ontologies specified in an ontology editor, such as Ontolingua, which has the ability to create CORBA IDL headers (OMG 1991) from ontology components. In this case, a CORBA IDL header is a skeleton of ontologies and its components, which should be complemented by implementation code written in Java. The second option is to generate Java interfaces from an ontology editor that has this capability. Since Ontolingua is not able to generate Java interfaces, we opted to develop an ontology editor to do this kind of work.

The remainder of this chapter introduces the ontology editor and the ontology browser. We show an example of how to query such a system and results get presented. The chapter's summary is in section 6.4.

6.1   The Ontology Editor

The ontology editor allows users to work on the specification of ontologies. After the ontology is specified, the user may query and update the ontologies using remote applications on the Internet.

The set of ontologies is represented in a hierarchy. The components of the hierarchy are classes modeled by their distinguishing features–*parts*, *functions*, and *attributes* (Figure 6-1). This structure for representing ontologies is extended from Rodríguez (2000) with the addition of *roles*. Roles allow for a richer representation of geographic entities and avoid the problems of multiple inheritance.
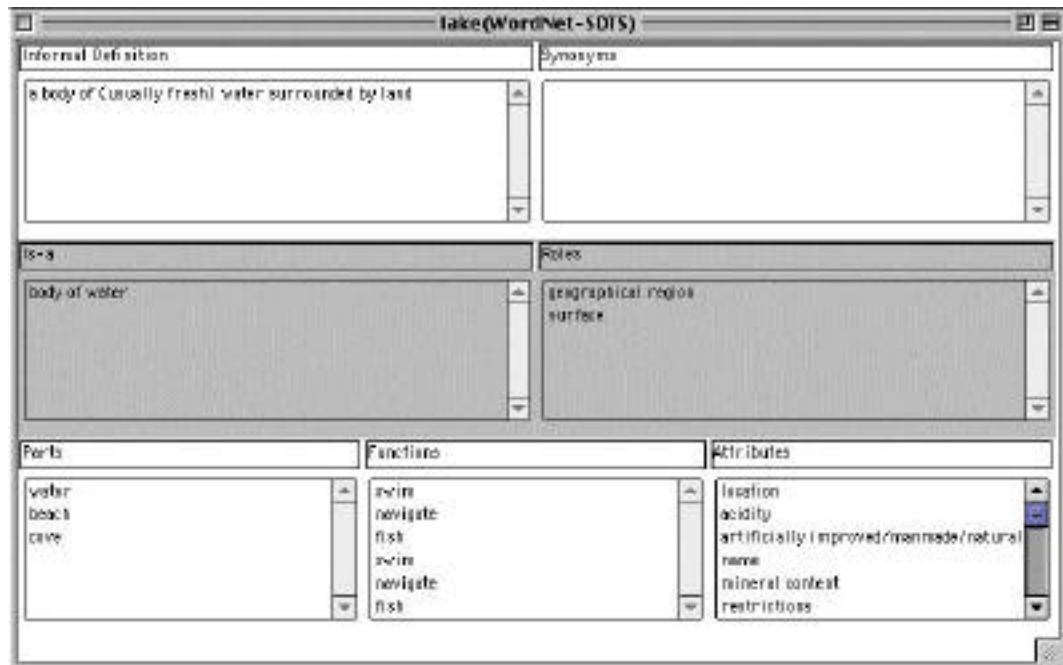
Figure 6-1 Basic structure on an ontology class.

Once the ontology is specified, the ontology editor has facilities for translating ontologies from repositories into application environments. We use Java as the implementation language. The basic mechanism for inheritance in Java is through the use of the keyword "extends". This mechanism allows a new class to inherit from only one parent class. The entities in the ontologies are translated into Java interfaces. A

Java interface describes the set of public methods that a class that implements the interface must support, and also their calling conventions. But a Java interface does not implement those methods. Each descendant class has to provide the code for each existing interface method (Figure 6-2).

```
Public interface lake

{

Vector roles;
//parts
Object water;
Object cove;
//functions
public void swim();
public void navigate();
public void fish();
//attributes
public String location;
public String acidity;
public String artificially_improved;
public String name;
public String mineral_content;
public String restrictions;
public String temperature;
public String charted_depth;
}
```

Figure 6-2 A Java interface for `lake`.

6.2 The Ontology Browser

In the ODGIS approach, the application program relies on classes derived from ontologies. These classes can be as simple as one entity or as complex as a part of an ontology . The application developer is able to browse the ontology that is the origin of these classes. The ontology browser has two important functions. First, it can be used during ontology specification by users who wish to collaborate in composing a shared ontology. Second, once the ontology has been specified, the browser is used to show the available geographic entities to the users. Mediators connect entities in ontologies to features in spatial databases.
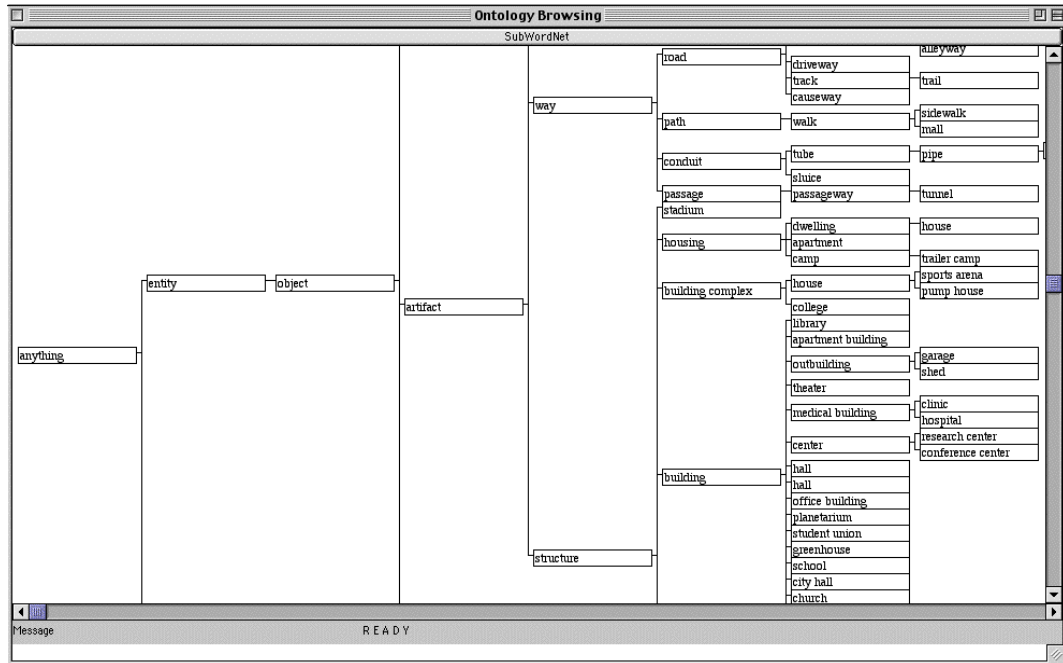
Figure 6-3 Browsing a top-level ontology.

For instance, a user wants to retrieve information about bodies of water for a determined region. First, the user browses the ontology server looking for the related classes. After that, the ontology server starts the mediators that look for the information and return a set of objects of the specified class. The results can be displayed (Figure 6-4) or can undergo any valid operation, such as statistical analysis.
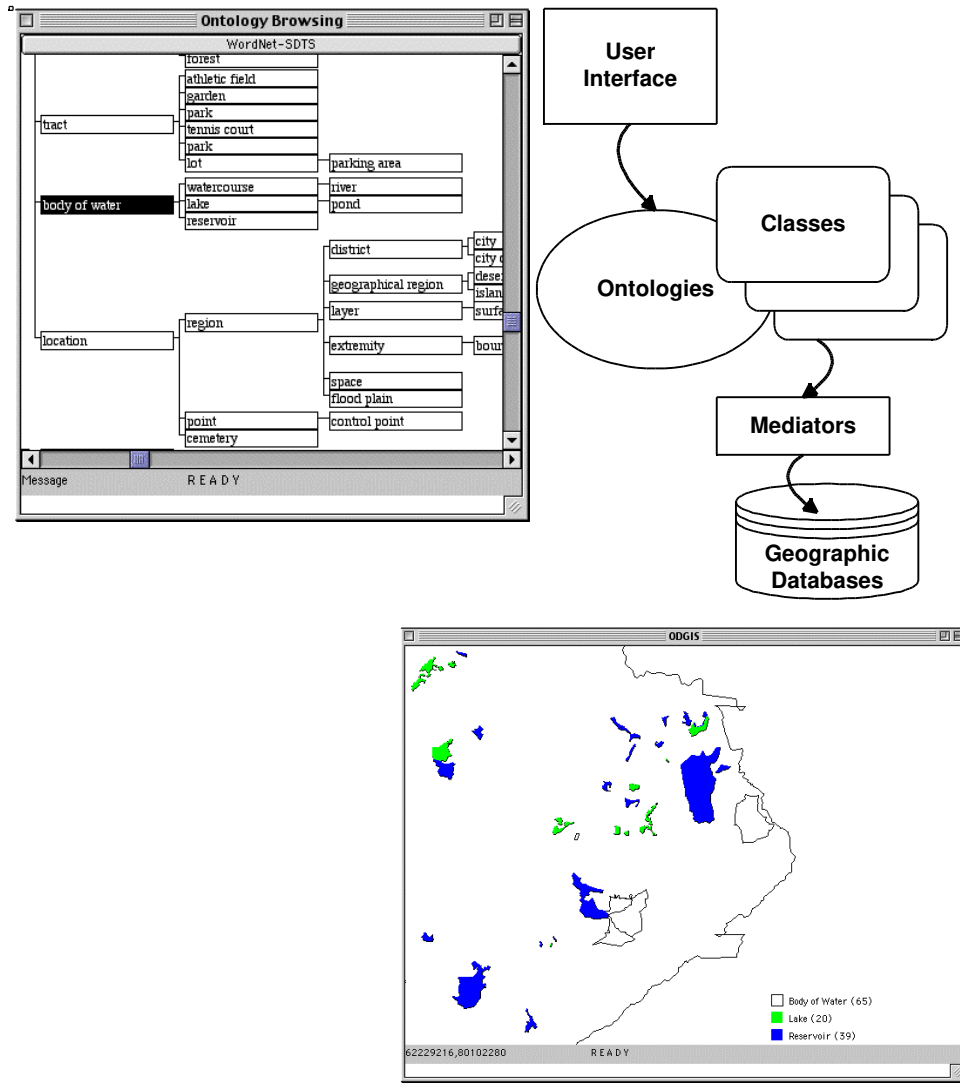
Figure 6-4 Schema for a query processing with an ODGIS.

6.3 Querying the System

The framework allows the user to browse at different levels of information. Ontologies are structured in a hierarchical way. This kind of organization leads to queries by level.

The entities chosen to be queried are `body of water`, `lake`, and `reservoir` (Figure 6-5).
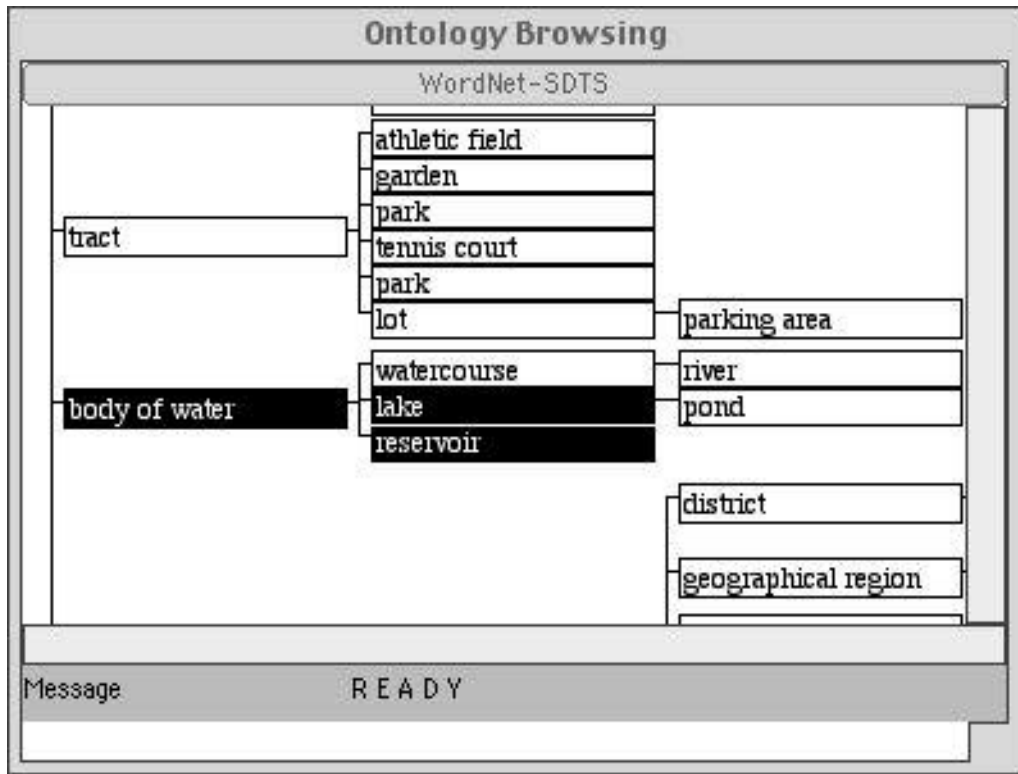
Figure 6-5 Query by level.

The user has to find the concepts in the ontology tree. The queries for `lake` presented the following result: 79 objects found (Figure 6-6). The query for `reservoir` is similar to the previous query. The result for `reservoir` was: 91 objects found (Figure 6-7).
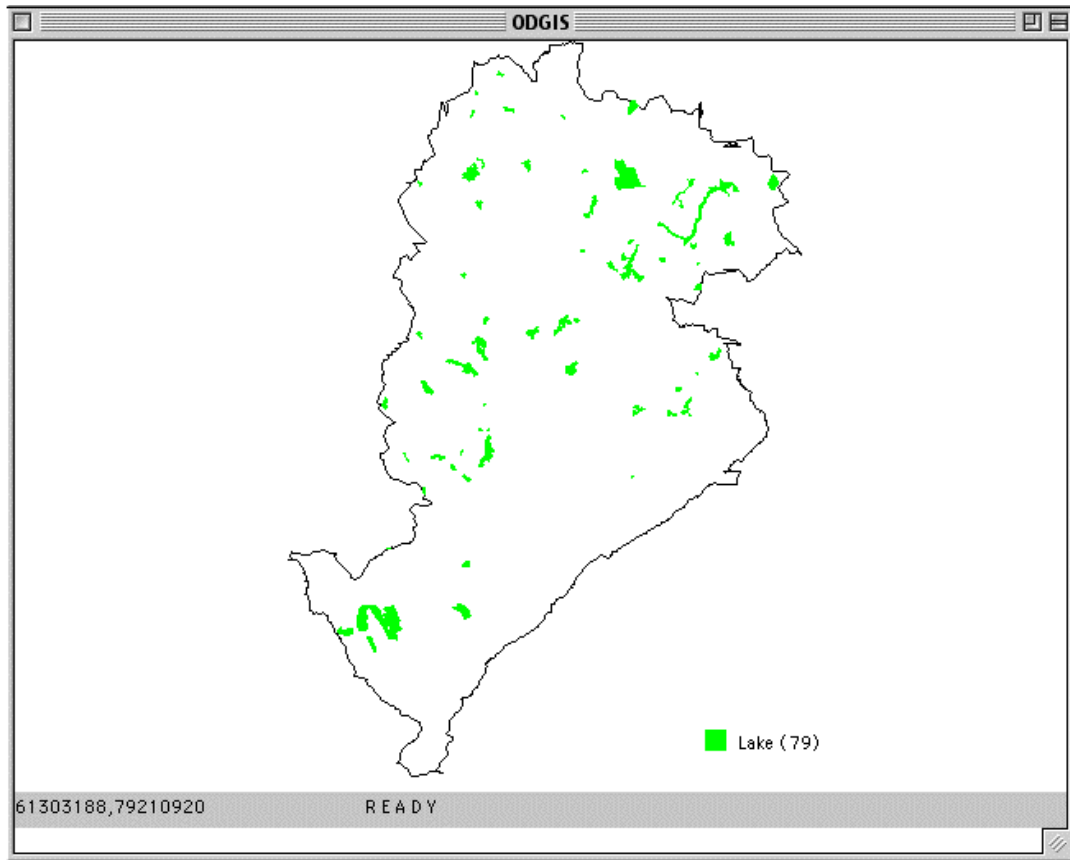
Figure 6-6 Query for `lake`.
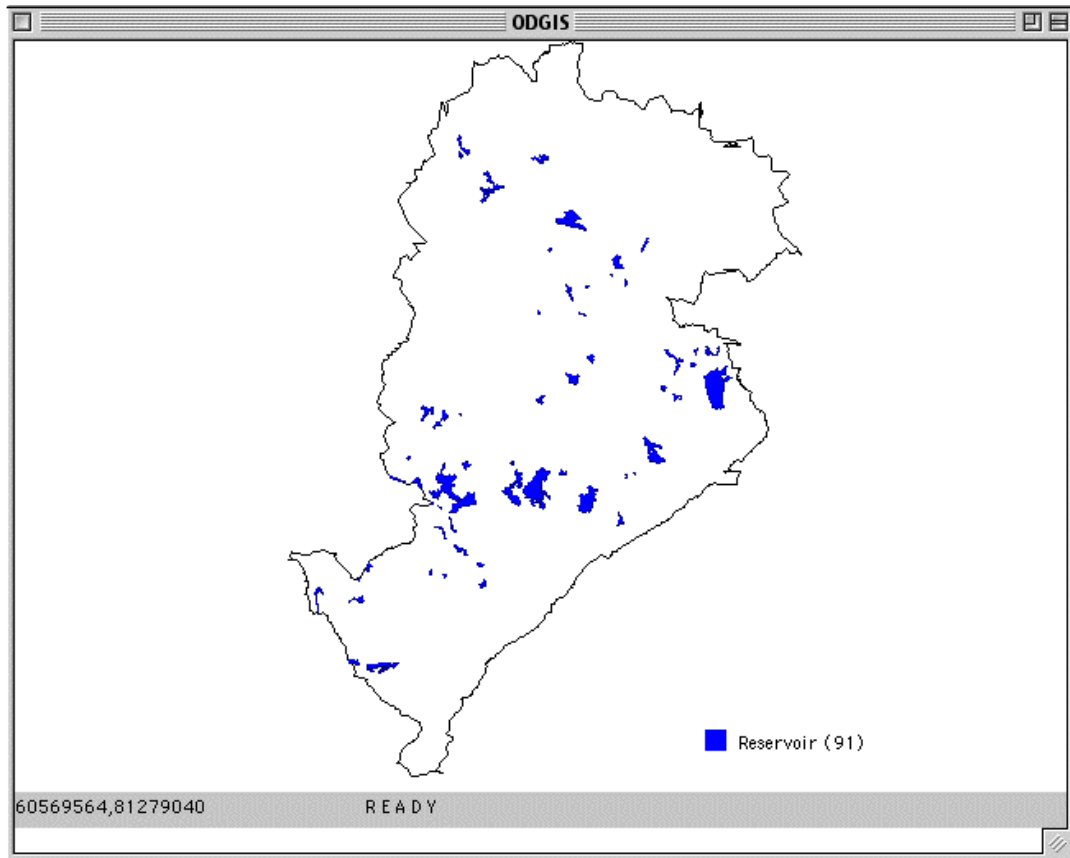
Figure 6-7 Query for `reservoir`.

When browsing the ontology of bodies of water, the user may choose to query for `body of water`. This entity is located one level higher than `lake` and `reservoir`, that is, it is necessary to explore the concept of `body of water`, finding that it includes both the concepts of `lake` and `reservoir`, thereby selecting both during the query. As a result, the query was performed at a higher semantic level. The result of the query for `body of water` was: 176 objects found (Figure 6-8).
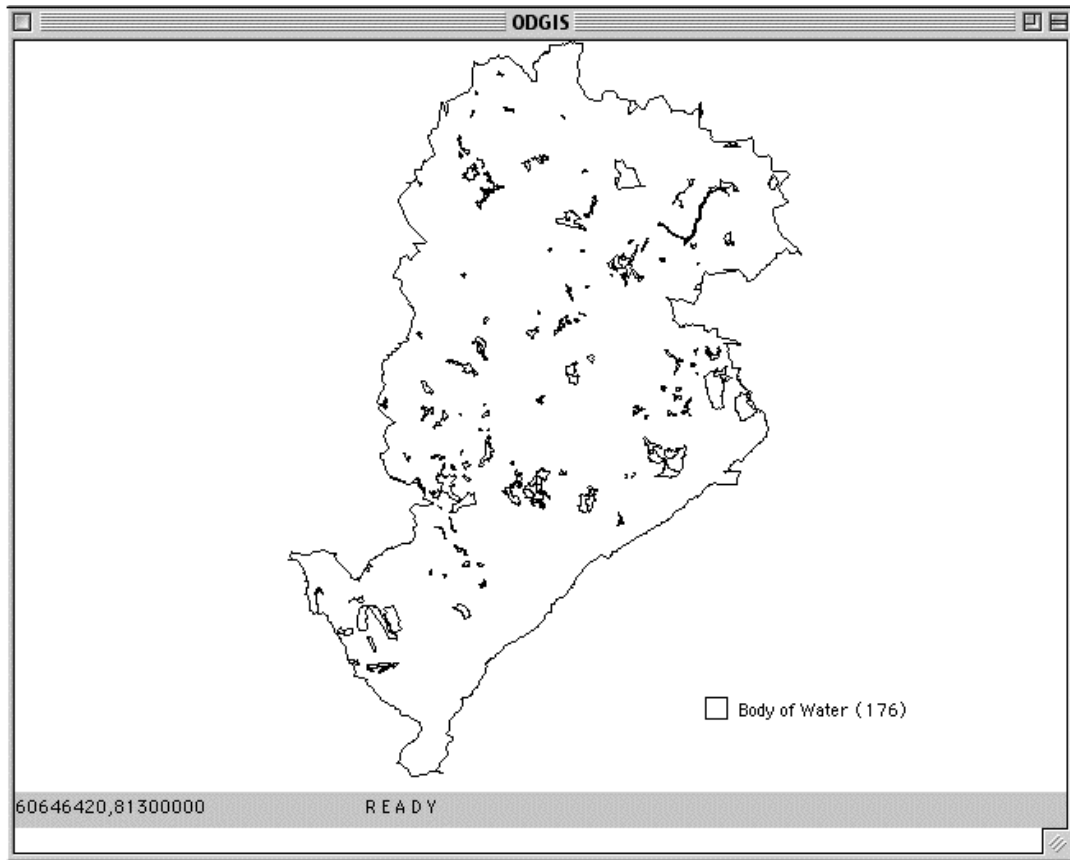
Figure 6-8 Query for `body of water`.

The results of the query using the semantic query for `body of water` can be compared against the results of the first two queries and against the result of the sum of these first two queries.

The query for `body of water` returned more objects (176) than the query for `lake` (79) and more than the query for `reservoir` (91). This result was expected and shows that with the semantic search broadened more adequate results are produced. These results correspond more closely to the user's notions about bodies of water, assuming that the concepts the user works with are adequately laid out as an ontology.

The expected results of a query for `body of water` could be the sum of the `lake` and `reservoir`, but we obtained a higher number (176) than that sum (170).

This result has two explanations. Both show the strength of the semantic approach for geographic information integration.

First, one reason for retrieving 176 objects instead of 170 is that since we are in a higher level in the hierarchy, other classes beyond `lake` and `reservoir` can be retrieved and classified as `body of water`, thus producing a broader result.

The second reason implies that, among the information systems integrated in this particular scenario of ODGIS, some of them can have information classified only at the higher conceptual level, i.e., `body of water`. The reasons for this more generic classification can be:

- Unclassified information collected from other sources.

- The source does not disclose the classification at a high level of detail. It only releases information at the lower semantic levels, because of security reasons or commercial purpose.

6.4   Summary

In this chapter we analyzed the implementation options for the main components of an ODGIS. We suggested the use of Java as an implementation language an ontology editor and an ontology browser was shown. The functionality of these systems components was demonstrated by a query for three different entities, performed at different levels of the ontology.

The next chapter of this thesis summarizes the work on ontology-driven geographic information systems and highlights the major findings of this thesis. Future directions of this research are also discussed.

Chapter 7

Conclusions and Future Work

This thesis focused on finding innovative ways to integrate geographic information. The thesis started the integration of information from entities of the physical universe. This approach differs from usual approaches that start from the implementation and representation universes. Our approach enables the integration of information based on its semantics content instead of dealing first with data formats and geometric representations. In order to integrate information it was necessary to integrate first ontologies. Therefore, this thesis studied new approaches to ontology integration. The question of how to measure the potential for information integration when combining two ontologies was also investigated here.

7.1   Summary of Thesis

This thesis investigated new ways to integrate geographic information. We chose to use ontologies as the foundation of the integration, because ontologies can represent real world entities using a sophisticated structure with components such as definitions, parts, functions, attributes, and rules of relationship. Furthermore, ontologies capture the semantics of information, can be represented in a formal language, and can be used to store related metadata. Ontologies can be used to establish agreements about diverse views of the world and consequently carry the meaning of the original ideas that are embedded in the representation of geographic phenomena in the human mind. The ontologies are linked to information sources through semantic mediators, therefore, the integration of ontologies leads to integration of information.

The integration of information depends on many factors, such as the way information is organized and the level of detail of each of its pieces. To face the problem of organization we proposed the use of ontologies as the basic representation

of geographic information. We chose a hierarchical organization, because hierarchies are a good way of representing the geographic world. Since geographic phenomena change over time and can also be seen as different things by different groups of people, we introduced the concept of roles. A geographic object can play different roles at the same time or during its lifetime depending on the point of view of a group of users. Roles act as the bridge between different levels of detail in an ontology structure. They are used also for networking ontologies of different domains.

The problem of the different levels of detail was approached by the introduction of a navigation mechanism that allows an object (i.e., the implementation of an ontology entity) to change its class by generalization or specialization. In a generalization, a more specific object drops some pieces of information and turns itself into a more general instance. In a specialization, a more general object gathers more information and becomes a more specific object. We also introduced the operation called role extraction, in which a role played by an object can be extracted and transformed into a new instance. This new instance acts as an independent object. Therefore, the new instance can be matched with an object associated with another entity in a different ontology.

We used a multiple-ontology approach to solve the problem of the different views of the world. The framework allows for the presence of multiple ontologies and provides mechanisms for integrating the ontologies. We also introduced a classification of ontologies of geographic phenomena in two types. One type is the phenomenological domain ontology (PDO). This ontology captures the different dimensions and internal properties of the geographic phenomena. This specific ontology is distinct and independent from the other type, the application domain ontology (ADO). This ontology is concerned with description of specific subjects and tasks that the GI scientists use as a source of information.

To validate our use of the concept of roles and hierarchies as the support for the ontology representation structure we made a simulation in which we measured the potential for information integration when combining two ontologies. Four different

evaluation measures were used to assess the potential for information integration: (1) using roles, (2) using roles and hierarchies, (3) using only hierarchies, and (4) using neither roles nor hierarchies. The use of a hierarchical structure improved the potential for information integration. So did the use of roles although to a much lesser extent than did the use of hierarchies. The combined effect of roles and hierarchies had a more positive effect in the potential for information integration than the use of roles alone or hierarchies alone. All three combinations gave better results than the results using neither roles nor hierarchies. The results supported the hypothesis that a model that incorporates hierarchies and roles has a potential to integrate more information than models that do not incorporate these concepts.

An ontology editor and an embedded translator from entities to classes were developed to support the knowledge-generation phase of the architecture. For the knowledge-use phase, a user interface to browse ontologies was also developed and the container of geographic objects was extended from Fonseca and Davis (1999).

7.2   Results and Major Findings

The main contribution of this thesis is the definition of a framework based on ontologies for the integration of geographic information. The use of ontologies allowed the integration of information to be based primarily on semantics contrasting thus with past approaches that were based on data formats and geometric representation. Our approach can be seen as complementary to existing approaches and it needs existing solutions to be fully implemented. The framework allows integration of information at different levels of detail. The general approach used by this thesis enables the use of the framework by developers of new GIS applications and by GIS database designers. The framework is up to date with requirements of modern information systems that should provide integration based on a semantic approach (Sheth 1999).

Another contribution of this thesis is that the framework provided a structure for the use and integration of multiple ontologies. Since it is difficult to find a unifying concept of space (Frank 1997) it is necessary to deal with multiple views of the

geographic world. GIS application developers need a tool to integrate multiple ontologies. The solution presented here allows for the integration of ontologies and hence the integration of information. The integration is accomplished through the combination of classes derived from diverse ontologies. This way it is possible to create geographic entities that are able to represent the complexity of the geographic world. The possibility of having multiple views of a single geographic object was supported by the use of hierarchies and roles in the representation of ontologies. Therefore, a geographic object can have more than one description. The support of multiple interpretations of the same geographic phenomenon answers the questions regarding different applications over the same area (Gahegan and Flack 1996).

The introduction of a mechanism to deal with changes of the level of detail proved to be useful, because information is available at different granularities and it also needs to be integrated at different levels of detail. The navigation mechanism allows an object to be transformed into a more general class or into a more specific class. We introduced also an operation called role extraction in which a role played by an object can be extracted and transformed in a new instance. This new instance acts as an independent object and can be matched with an object associated with another entity in a different ontology.

A major result of this thesis is a new methodology to measure the potential for information integration when combining two ontologies. Based on the structure used to represent ontologies, the measurement considered each one of the possible matches: entity vs. entity, entity vs. role, and role vs. role, entity vs. parent of an entity, and role vs parent of an entity. The measurement was used in an experiment that evaluated the influence of the use of hierarchical structure and roles for representing ontologies on the potential for information integration. The use of a hierarchical structure improved the potential for information integration. The use of roles also presented good results for the potential for information integration although to a much lesser extent than the use of hierarchies did. The combined effect of roles and hierarchies had a more positive effect in the potential for information integration than the use of roles alone or hierarchies alone. All three combinations gave better results than the results using

neither roles nor hierarchies. The results of the experiment supported the hypothesis of this thesis that a model that incorporates hierarchies and roles has a potential to integrate more information than models that do not incorporate these concepts.

This thesis presented a model for the integration of ontologies that is flexible enough to accommodate two different perspectives of ontologies. The first perspective is that there is one Ontology and that we can reach a consensus about it through the refinement of the concepts step by step over time. The other perspective does not accept this one Ontology and says that it is necessary to live with incompatible views of reality. The model presented here is based on the assumption that this one Ontology exists, at least inside small communities. Small here can vary from a group of five or six people in an office to 100 people in a local government department. We argue that there is consensus inside each community about the geographic phenomena that are part of the basic domain of this group. Using this model we can start combining ontologies at a higher level of abstraction and this way composing new and more comprehensive ontologies. There will be always some amount of information lost when combining different ontologies, but at the same time there is always some amount that remains available after the integration. This way we can refine progressively groups of ontologies and maybe one day reach this one big Ontology. The answer to the question if this is possible or not is subject to further study.

7.3   Future Work

A variety of issues remain to be resolved. One of them is to study the effect of other ontology components, besides roles and hierarchies, on the potential for information integration when combining two ontologies. The use of ontologies in information systems is an emerging field and many questions are open. In the next sections we discuss new questions that became apparent through the results of this thesis. They address ontology integration, geographic information retrieval on the web, ontology specification, ontology of actions, and ontology of images.

7.3.1 Other Approaches to Ontology Integration

Building an ODGIS requires an ontological commitment from users and information providers. User associations can be used as anchor points to start the production of ontologies. A further study should investigate how to incorporate approaches that allow composition of pre-existing, independently developed ontologies, for instance, through the use of a context algebra to compose diverse ontologies (Wiederhold and Jannink 1998). The solution presented here specifies ontologies through the use of parts, functions and attributes. The use of an algebraic definition of semantics based on operations (Kuhn 1994) and the matching of synonym, hyponym, and hypernym terms (Kashyap and Sheth 1996; Mena *et al.* 1996; Mena *et al.* 1998) in the integration of ontologies should be studied.

7.3.2 Ontologies for the Web

The commercial structure that the Internet is bringing is quite different from the past. One of the models that is being established gives for free basic services and charges for more sophisticated solutions. This willingness to offer products to attract new customers can be the foundation for future ontology-driven information systems. Top-level ontologies may be offered for free to users. The more elaborate ontologies–domain, task, and application ontologies–can be charged and customized according to the users needs. So, one of the future directions of this research is to analyze the role of these service providers as the focal points for ontology servers.

The integration of informal ontologies representing the information available today on the Internet was beyond the scope of this thesis, however, the dynamic object-oriented approach suggested here can be extended to fulfill the requirements of future geographic information systems that are strongly based on the Internet and on the integration of diverse sources of information.

7.3.3 Foundations of Ontology Specification

Frank (1999) points out that the formal specification of spatial objects and spatial relations is one of the first steps for GIS interoperability. This specification should be

close to what people use in their everyday lives (Egenhofer and Mark 1995). The foundation for the specification of large-scale objects using image schemata has been laid (Rodríguez and Egenhofer 1997; Egenhofer and Rodríguez 1999; Frank and Raubal 1999; Rodríguez and Egenhofer 2000). One interesting line of research is to explore how image schemata can be used as the support for the specification of geo-ontologies.

7.3.4   Action-Driven Ontologies

The main emphasis of current approaches to geo-ontologies pursue a static perspective of geographic reality. In this thesis we followed this direction considering only the operations from a high abstraction level and thus only operations inherent to the existing geographic objects are modeled. Future work should consider the study of ontologies of the operations that can be performed by a GIS user. These ontologies should capture the intended use of the information and consider an ontology of actions independent of the existing objects. New approaches to geo-ontologies can be based on a definition of *space* proposed Santos (1997) "Space is a system of objects and a system of actions." In this view, space consists of natural and technical objects, and the actions that have transformed nature by human decision. We cannot capture the full meaning of an object, without taking into account the intentionality of the human action that has produced the object and placed it in a location on space. These kind of ontologies are called action-driven, considering that they should capture, to some extent, knowledge domains such as cartographic modeling, spatial queries, spatial statistics, cellular automata and dynamic modeling. Action-driven ontologies should consist of three different components:

- A description of a set of entities, using concepts from the user domain and their relation to geometrical representations in a computer.

- A description of a set of actions, including both the knowledge domain vocabulary and its relation to the GIS operations and to the data that are being produced.

- A description of the intended use of such information, including the intermediate and final products.

7.3.5   Ontology of Images

Remotely sensed imagery is one of the most frequently used sources of spatial data currently available to researchers. There is a large variety of spatial and spectral resolutions for remote sensing images, ranging from IKONOS 1-meter panchromatic images to the polarimetric radar images soon to be part of the next generation of RADARSAT and JERS satellites.

In spite of the ubiquitous nature of remote sensing imagery, and more than 30 years experience of data gathering, processing, and analysis, their ontological status is still open to debate. It is surprisingly difficult to provide a straightforward answer to a very basic question: What is in an image? We can ask the same question in a different way: What is the ontological status of the information content of remote sensing imagery? The answer to this question requires the specification of the phenomenological domain ontology. This line of research should examine different possible answers to this question generating a set of ontologies that will allow a more complete understanding of the role of images as sources of geographic information.

Although these topics are beyond of the scope of this thesis, they give directions for future work. The framework presented here is a foundation for future ontology-driven information systems. The highly semantic approach used here follows a trend of modern information systems. Although the development of the large amount of ontologies necessary to fully unlash the power of these kind of systems is still an ongoing effort, users of future information systems will be able to deal with information in a much more intuitive and easy way than in today's keyword based systems.

References

Abel, D., Ooi, B., Tan, K.-L., and Tan, S. (1998) Towards Integrated Geographical Information Processing. *International Journal of Geographical Information Science* 12(4): 353-371.

Albano, A., Bergamini, R., Ghelli, G., and Orsini, R. (1993) An Object Data Model with Roles. in: R. Agrawal, S. Baker, and D. Bell, (Eds.), *19th International Conference on Very Large Data Bases*, Dublin, Ireland, pp. 39-51.

Arctur, D., Hair, D., Timson, G., Martin, E., and Fegeas, R. (1998) Issues and Prospects for the Next Generation of the Spatial Data Transfer Standard (SDTS). *International Journal of Geographical information Science* 12(4): 403-425.

Batini, C., Lenzerini, M. and Navathe, S. (1986) A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys* 18(4): 323-364.

Beard, K. (1987) How To Survive A Single Detailed Database. in: N. R. Chrisman, (Ed.), *AUTO-CARTO 8, Eighth International Symposium on Computer-Assisted Cartography*, Baltimore, MD, pp. 211-220.

Bergamaschi, S., Castano, S., Vermercati, S., Montanari, S., and Vincini, M. (1998) An Intelligent Approach to Information Integration. in: N. Guarino, (Ed.), *Formal Ontology in Information Systems*. IOS Press, Amsterdan, Netherlands, pp. 253-268.

Bertolotto, M. and Egenhofer, M. (1999) Progressive Vector Transmission. in: C. B. Medeiros, (Ed.), *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, pp. 152-157.

Betz, M. (1994) Interoperable Objects. *Dr. Dobb's Journal* 4(220): 22-26.

Bishr, Y. (1997) *Semantic Aspects of Interoperable GIS*. Ph.D. Thesis, Wageningen Agricultural University, The Netherlands.

Bishr, Y. (1998) Overcoming the Semantic and Other Barriers to GIS Interoperability. *International Journal of Geographical Information Science* 12(4): 299-314.

Bock, C. and Odell, J. (1998) A More Complete Model of Relations and Their Implementation: Roles. *Journal of Object-Oriented Programming* 11(2): 51-54.

Bryant, D. and Tversky, B. (1992) Internal and External Spatial Frameworks for Representing Described Scenes. *Journal of Memory and Language* 31: 74-98.

Burrough, P. and Frank, A., (Eds.), (1996) *Spatial Conceptual Models for Geographic Objects with Undetermined Boundaries*. Taylor & Francis, London.

Câmara, G., Souza, R., Freitas, U., and Monteiro, A. (1999) Interoperability in Practice: Problems in Semantic Conversion from Current Technology to OpenGIS. in: A. Vckovski, K. Brassel, and H.-J. Schek, (Eds.), *Interoperating Geographic Information Systems - Second International Conference, INTEROP'99. Lecture Notes in Computer Science* 1580, Springer-Verlag, Berlin, pp. 129-138.

Câmara, G., Souza, R. C. M., Freitas, U. M., and Garrido, J. C. P. (1996) SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. *Computers and Graphics* 20(3): 395-403.

Car, A. and Frank, A. (1994) General Principles of Hierarchical Spatial Reasoning–The Case of Wayfinding. in: T. Waugh and R. Healey, (Eds.), *Sixth International Symposium on Spatial Data Handling*, Edinburgh, Scotland, pp. 646-664.

Cardelli, L. (1984) A Semantics of Multiple Inheritance. in: G. Kahn, D. McQueen, and G. Plotkin, (Eds.), *Semantics of Data Types. Lecture Notes in Computer Science* 173, Springer-Verlag, New York, pp. 51-67.

Clemens, P. (1996) *Coming Attractions in Software Archictecture*. Carnegie Mellon University, Pittsburgh, PA, Technical Report CMU/SEI-96-TR-008.

Clocksin, W. and Mellish, C. (1981) *Programming in Prolog*. Springer-Verlag, New York.

Couclelis, H. (1992) People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS. in: A. U. Frank, I. Campari, and U. Formentini, (Eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science* 639, Springer-Verlag, New York, pp. 65-77.

Davis, C. and Laender, A. (1999) Multiple Representations in GIS: Materialization Through Map Generalization, Geometric and Spatial Analysis Operations. in: C. B. Medeiros, (Ed.), *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, pp. 60-65.

Egenhofer, M. and Frank, A. (1992) Object-Oriented Modeling for GIS. *Journal of the Urban and Regional Information Systems Association* 4(2): 3-19.

Egenhofer, M. and Mark, D. (1995) Naive Geography. in: A. Frank and W. Kuhn, (Eds.), *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95. Lecture Notes in Computer Science* 988, Springer-Verlag, Berlin, pp. 1-15.

Egenhofer, M. and Rodríguez, A. (1999) Relation Algebras over Containers and Surfaces: An Ontological Study of a Room Space. *Spatial Cognition and Computation* 1(2): 150-180.

Elmagarmid, A. and Pu, C. (1990) Guest Editors' Introduction to the Special Issue on Heterogeneous Databases. *ACM Computing Surveys* 22(3): 175-178.

Farquhar, A., Fikes, R. and Rice, J. (1996) *The Ontolingua Server: A Tool for Collaborative Ontology Construction*. Knowledge Systems Laboratory, Stanford University, Stanford, CA, Technical Report KSL 96-26.

Fonseca, F. and Davis, C. (1999) Using the Internet to Access Geographic Information: An OpenGis Prototype. in: M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, (Eds.), *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Norwell, MA, pp. 313-324.

Fonseca, F. and Egenhofer, M. (1999) Ontology-Driven Geographic Information Systems. in: C. B. Medeiros, (Ed.), *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, pp. 14-19.

Fonseca, F., Egenhofer, M. and Davis, C. (2000) Ontology-Driven Information Integration. in: C. Bettini and A. Montanari, (Eds.), *The AAAI—2000 Workshop on Spatial and Temporal Granularity*, Austin, TX, pp. 61-64.

Frank, A. (1997) Spatial Ontology. in: O. Stock, (Ed.), *Spatial and Temporal Reasoning*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 135-153.

Frank, A. and Raubal, M. (1999) Formal Specification of Image Schemata—a Step towards Interoperability in Geographic Information Systems. *Spatial Cognition and Computation* 1(1): 67-101.

Gahegan, M. (1999) Characterizing the Semantic Content of Geographic Data, Models, and Systems. in: M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, (Eds.), *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Norwell, MA, pp. 71-84.

Gahegan, M. and Flack, J. (1996) A Model to Support the Integration of Image Understanding Techniques within a GIS. *Photogrammetric Engineering & Remote Sensing* 62(5): 483-490.

Genesereth, M. R. (1990) *The Epikit Manual*. Epistemics, Palo Alto, CA, Technical Report.

Genesereth, M. R. and Fikes, R. E. (1992) *Knowledge Interchange Format*. Stanford University, Knowledge Systems Laboratory, Technical Report KSL-92-86.

Gomes, J. and Velho, L. (1995) Abstraction Paradigms for Computer Graphics. *The Visual Computer* 11(5): 227-239.

Goodchild, M. (1992) Geographical Data Modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M., Egenhofer, M., Fegeas, R., and Kottman, C. (1999a) *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Norwell, MA.

Goodchild, M., Egenhofer, M., Kemp, K., Mark, D., and Sheppard, E. (1999b) Introduction to the Varenius Project. *International Journal of Geographical Information Science* 13(8): 731-745.

Gruber, T. (1992) *A Translation Approach to Portable Ontology Specifications*. Knowledge Systems Laboratory, Stanford University, Stanford, CA, Technical Report KSL 92-71.

Guarino, N. (1992) Concepts, Attributes and Arbitrary Relations. *Data and Knowledge Engineering* 8: 249-261.

Guarino, N. (1997) Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. in: M. Pazienza, (Ed.), *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, International Summer School, SCIE-97*, Frascati, Italy, pp. 139-170.

Guarino, N. (1998) Formal Ontology and Information Systems. in: N. Guarino, (Ed.), *Formal Ontology in Information Systems*. IOS Press, Amsterdam, Netherlands, pp. 3-15.

Guarino, N. and Welty, C. (2000a) A Formal Ontology of Properties. in: R. Dieng and O. Corby, (Eds.), *Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management. Lecture Notes in Computer Science* 1920, 97-112.

Guarino, N. and Welty, C. (2000b) Ontological Analysis of Taxonomic Relationships. in: A. Laender and V. Storey, (Eds.), *Proceedings of ER-2000: The 19th International Conference on Conceptual Modeling. Lecture Notes in Computer Science* 1920, 210-224.

Halbert, D. C. and O'Brien, P. D. (1987) Using Types and Inheritance in Object-Oriented Languages. in: J. Bézivin, J.-M. Hullot, P. Cointe, and H. Lieberman, (Eds.), *Proceedings of ECOOP'87 European Conference on Object-Oriented Programming. Lecture Notes in Computer Science* 276, 20-31.

Harvey, F. (1999) Designing for Interoperability: Overcoming Semantic Differences. in: M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, (Eds.), *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Norwell, MA, pp. 85-98.

Hornsby, K. (1999) *Identity-Based Reasoning about Spatio-Temporal Change*. Ph.D. Thesis, University of Maine, Orono.

Huxhold, W. and Levinsohn, A. (1995) *Managing Geographic Information System Projects*. Oxford University Press, New York.

Huxhold, W. E. (1991) *An Introduction to Urban Geographic Information Systems*. Oxford University Press, New York.

Kashyap, V. and Sheth, A. (1996) Semantic Heterogeneity in Global Information System: The Role of Metadata, Context and Ontologies. in: M. Papazoglou and G. Schlageter, (Eds.), *Cooperative Information Systems: Current Trends and Directions*. Academic Press, London, pp. 139-178.

Kent, W. (1993) Object Orientation and Interoperability. in: *Advances in Object-Oriented Database Systems*. *NATO Advanced Study Institute on Object-Oriented Database Systems* 130, Springer, Izmir, Kusadasi, Turkey, pp. 287-305.

Kuhn, W. (1991) Are Displays Maps or Views? in: D. Mark and D. White, (Eds.), *AUTO-CARTO 10, Tenth International Symposium on Computer-Assisted Cartography*, Baltimore, MD, pp. 261-274.

Kuhn, W. (1994) Defining Semantics for Spatial Data Transfer. in: T. Waugh and R. Healey, (Eds.), *Sixth International Symposium on Spatial Data Handling*, Edinburgh, Scotland, pp. 973-987.

Kuno, H. A. and Rundensteiner, E. A. (1996) The MultiView OODB View System: Design and Implementation. *TAPOS - Theory and Practice of Object Systems* 2(3): 202-225.

Langacker, R. W. (1987) *Foundations of Cognitive Grammar*. Stanford University Press, Stanford, CA.

Lewandowsky, S. M. (1998) Frameworks for Component-Based Client/Server Computing. *ACM Computing Surveys* 30(1): 4-27.

Mark, D. (1993) Toward a Theoretical Framework for Geographic Entity Types. in: A. Frank and I. Campari, (Eds.), *Spatial Information Theory. Lectures Notes in Computer Science* 716, Springer-Verlag, Berlin, pp. 270-283.

Marr, D. (1982) *Vision*. W. H. Freeman, New York.

McKee, L. and Buehler, K., (Eds.), (1996) *The Open GIS Guide*. Open GIS Consortium, Inc, Wayland, MA.

Mena, E., Kashyap, V., Illarramendi, A., and Sheth, A. (1998) Domain Specific Ontologies for Semantic Information Brokering on the Global Information Infrastructure. in: N. Guarino, (Ed.), *Formal Ontology in Information Systems*. IOS Press, Amsterdam, pp. 269-283.

Mena, E., Kashyap, V., Sheth, A., and Illarramendi, A. (1996) OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies. in: *First IFCIS International Conference on Cooperative Information Systems (CoopIS'96)*, Brussels, Belgium, pp. 14-25.

Mendelzon, A. O., Milo, T. and Waller, E. (1994) Object Migration. in: *Proceedings of the 13th ACM SIGACT—SIGMOD—SIGART Symposium on Principles of Database Systems*, Minneapolis, MN, pp. 232-242.

Meyer, B. (1988) *Object-Oriented Software Construction*. Prentice Hall, New York.

Miller, G. A. (1995) WordNet: A Lexical Database for English. *Communications of the ACM* 38(11): 39-41.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. (1991) Enabling Technology for Knowledge Sharing. *AI Magazine* 12(3): 13-36.

Nunes, J. (1991) Geographic Space as a Set of Concrete Geographical Entities. in: D. Mark and A. Frank, (Eds.), *Cognitive and Linguistic Aspects of Geographic Space*. Kluwer Academic Publishers, Norwell, MA, pp. 9-33.

OMG, (Ed.), (1991) *The Common Object Request Broker: Architecture and Specification, Revision1.1*. OMG Document No. 91.12.1, Framingham, MA.

Papakonstantinou, Y., Garcia-Molina, H. and Widom, J. (1995) Object Exchange Across Heterogeneous Information Sources. in: *IEEE International Conference on Data Engineering*, Taipei, Taiwan, pp. 251-260.

Parent, C., Spaccapietra, S. and Zimanyi, E. (2000) MurMur: Database Management of Multiple Representations. in: C. Bettini and A. Montanari, (Eds.), *The AAAI-2000 Workshop on Spatial and Temporal Granularity*, Austin, TX, pp. 61-64.

Pernici, B. (1990) Objects with Roles. in: *IEEE/ACM Conference on Office Information Systems*, Cambridge, MA, pp. 205-215.

Requicha, A. (1980) Representations for Rigid Solids: Theory Methods and Systems. *ACM Computing Surveys* 12(4): 437-464.

Rodríguez, A. (2000) *Assessing Semantic Similarity among Spatial Entity Classes*. Ph.D. Thesis, University of Maine, Orono, ME.

Rodríguez, A. and Egenhofer, M. (1997) Image-Schemata-Based Spatial Inferences: The Container-Surface Algebra. in: S. Hirtle and A. Frank, (Eds.), *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '97. Lecture Notes in Computer Science* 1329, Springer Verlag, Berlin, pp. 35-52.

Rodríguez, A. and Egenhofer, M. (2000) A Comparison of Inferences about Containers and Surfaces in Small-Scale and Large-Scale Spaces. *Journal of Visual Languages and Computing* 11(6): 639-662.

Rodríguez, A., Egenhofer, M. and Rugg, R. (1999) Assessing Semantic Similarity Among Geospatial Feature Class Definitions. in: A. Vckovski, K. Brassel, and H.-J. Schek, (Eds.), *Interoperating Geographic Information Systems—Second International Conference, INTEROP'99. Lecture Notes in Computer Science* 1580, Springer-Verlag, Berlin, pp. 1-16.

Santos, M. (1997) *A Natureza do Espaço*. Hucitec, São Paulo, Brazil.

Sheth, A. (1999) Changing Focus on Interoperability in Information Systems: from System, Syntax, structure to Semantics. in: M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, (Eds.), *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Norwell, MA, pp. 5-29.

Sheth, A. and Larson, J. (1990) Federated Databases Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys* 22(3): 183-236.

Shimabukuro, Y. E., Batista, G. T., Mello, E. M. K., Moreira, J. C., and Duarte, V. (1998) Using Shade Fraction Image Segmentation to Evaluate Deforestation in Landsat Thematic Mapper Images of the Amazon Region. *International Journal of Remote Sensing* 19(3): 535-541.

Smith, B. (1995) On Drawing Lines on a Map. in: A. Frank and W. Kuhn, (Eds.), *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '95. Lecture Notes in Computer Science* 988, Springer Verlag, Berlin, pp. 475-484.

Smith, B. (1998) An Introduction to Ontology. in: D. Peuquet, B. Smith, and B. Brogaard, (Eds.), *The Ontology of Fields*. National Center for Geographic Information and Analysis, Santa Barbara, CA, pp. 10-14.

Smith, B. and Mark, D. (1998) Ontology and Geographic Kinds. in: *International Symposium on Spatial Data Handling*, Vancouver, BC, Canada, pp. 308-320.

Soley, R. M. and Kent, W. (1995) The OMG Object Model. in: W. Kim, (Ed.), *Modern Database Systems: the Object Model, Interoperability and Beyond*. Addison-Wesley Publishing Company, New York, pp. 18-41.

Sondheim, M., Gardels, K. and Buehler, K. (1999) GIS Interoperability. in: P. Longley, M. Goodchild, D. Maguire, and D. Rhind, (Eds.), *Geographical Information Systems 1 Principles and Technical Issues*. John Wiley & Sons, New York, pp. 347-358.

Steimann, F. (2000) On the Representation of Roles in Object-Oriented and Conceptual Modelling. *Data & Knowledge Engineering* 35(1): 83-106.

Steimann, F. (2001) Roles = Interfaces: A Merger of Concepts. *Journal of Object-Oriented Programming* (in press).

Stell, J. and Worboys, M. (1998) Stratified Map Spaces: A Formal Basis for Multi-resolution Spatial Databases. in: *International Symposium on Spatial Data Handling*, Vancouver, BC, Canada, pp. 180-189.

Su, J. (1991) Dynamic Constraints and Object Migration. in: G. Lohman, A. Sernadas, and R. Camps, (Eds.), *17th International Conference on Very Large Data Bases*, Barcelona, Spain, pp. 233-242.

Tanaka, M. and Ichikawa, T. (1988) A Visual User Interface for Map Information Retrieval Based on Semantic Significance. *IEEE Transactions on Software Engineering* SE 14(5): 666-670.

Tempero, E. and Biddle, R. (1998) *Simulating Multiple Inheritance in Java*. Victoria University of Wellington, School of Mathematical and Computing Sciences, Wellington, New Zealand, Technical Report CS-TR-98/1.

Timpf, S. and Frank, A. (1997) Using Hierarchical Spatial Data Structures for Hierarchical Spatial Reasoning. in: S. Hirtle and A. Frank, (Eds.), *Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '97. Lecture Notes in Computer Science* 1329, Springer-Verlag, Berlin, pp. 69-83.

USGS (1998), *View of the Spatial Data Transfer Standard (SDTS) Document*, http://mcmcweb.er.usgs.gov/sdts/standard.html.

Vckovski, A. (1997) *Interoperable and Distributed Geoprocessing*. Ph.D. Thesis, Universität Zürich, Zurich, Switzerland.

Volta, G. and Egenhofer, M. (1993) Interaction with GIS Attribute Data Based on Categorical Coverages. in: A. Frank and I. Campari, (Eds.), *Spatial Information Theory, European Conference COSIT '93. Lecture Notes in Computer Science* 716, Springer-Verlag, New York, pp. 215-233.

Wiederhold, G. (1991) *Mediators in the Architecture of Future Information Systems*. Stanford University, Technical Report.

Wiederhold, G. (1994) Interoperation, Mediation and Ontologies. in: *International Symposium on Fifth Generation Computer Systems (FGCS94)*, Tokyo, Japan, pp. 33-48.

Wiederhold, G. (1998) *Value-added Middleware: Mediators*. Stanford University, Technical Report.

Wiederhold, G. (1999) Mediation to Deal with Heterogeneous Data Sources. in: A. Vckovski, K. Brassel, and H.-J. Schek, (Eds.), *Interoperating Geographic Information Systems—Second International Conference, INTEROP'99*. *Lecture Notes in Computer Science* 1580, Springer-Verlag, Berlin, pp. 1-16.

Wiederhold, G. and Jannink, J. (1998) *Composing Diverse Ontologies*. Stanford University, Technical Report.

Wong, R., Chau, H. and Lochovsky, F. (1997) A Data Model and Semantics of Objects with Dynamic Roles. in: M. Jackson and C. Pu, (Eds.), *13th International Conference on Data Engineering*, Birmingham, UK, pp. 402-411.

Worboys, M. and Deen, S. (1991) Semantic Heterogeneity in Geographic Databases. *SIGMOD RECORD* 20(4): 30-34.

Biography of the Author

Frederico Torres Fonseca was born in Belo Horizonte, Brazil, on November 27, 1955. He graduated from the Universidade Federal de Minas Gerais in 1977 with a degree in Data Processing. He also graduated from the Universidade Católica de Minas Gerais in 1978 with a degree in Mechanical Engineering. Since his graduation, Frederico has worked as a system analyst at several companies in Brazil and lately as GIS analyst at the Prodabel Belo Horizonte, Brazil. In 1995 he started his Master degree in Public Administration and Computer Science at Fundação João Pinheiro in Brazil. Since 1998 Frederico has been a graduate research assistant at the National Center for Geographic Information and Analysis. He is the recipient of the 1999 ESRI/IGIF Scholarship, a NASA/EPSCoR fellowship, and the recipient of the 2000 Graduate Research Assistant Award of the College of Engineering of The University of Maine. Frederico is a candidate for the Doctor of Philosophy degree in Spatial Information Science and Engineering from The University of Maine in May, 2001.