Role-Based Geographic Information Integration

Frederico Torres Fonseca
School of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802
fredfonseca@ist.psu.edu

Abstract. The integration of geographic 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. 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.

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 work. This integration is usually done through an interface that acts as the integrator of information originating from different places.

Integration of geographic information (Vckovski 1998; Goodchild *et al.* 1999) 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). 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 OpenGIS data model deals primarily with representations of geographic information. 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. Therefore, a common format alone is not enough to provide information integration based on meaning (Mark 1993). 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. 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 work 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. 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 (Fonseca 2001). 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 (Guarino 1998; Smith and Mark 1998). This work 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 (Fonseca *et al.* 2001).

2 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 work are "how can these ontologies be combined, leading to information integration?" And, "what are the mechanisms for change of levels inside ontologies?"

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.

3 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 (2000). The concept of role as interfaces as we use in the implementation of this work 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 work 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 work 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 2000) 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).

4 Measuring the Integration of Ontologies

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 (1) one entity in the first ontology with one entity in the second ontology (E-E); (2) one entity in the first ontology with one role in the second ontology (E-R); (3) one role in the first ontology with one entity in the second ontology (R-E); (4) one role in the first ontology with one role in the second ontology (R-R).

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 (1) one entity in the first ontology with the parent of one entity in the second ontology ($E\text{-}P_E$); (2) one role in the first ontology with the parent of one entity in the second ontology ($R\text{-}P_E$). The final result reflects the sum of the amount integrated in all the matches of E-R, E-E, R-R, R-E, $E\text{-}P_E$, and $R\text{-}P_E$.

5 A Method for Evaluating the Potential for Integration of Information

A domain can be represented as a set of ontologies (Equation 5.1). Domain D : $\{O_n\}$ $n \geq 1$ (5.1)

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

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

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

Entity E: {id, R} (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: $\{E_i\}$ $E_i \subseteq D$, $0 \leq i \leq n$, where n is the size of the set. (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.

$I = \sum \text{compare}_k (E_1,E_2)$ $E_1 \subseteq O_n$, $E_2 \subseteq O_m$, where k is n x m (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 remainder of this section 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.

5.1 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.

$I_R = E_E + R_E + R_R$ (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. $I = \sum (I_R)_k$, where k is n x m. (5.7)

5.2 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:

$I_{R+H} = I_R + I_H$ (5.8)

The expanded formula is given in Equation 5.9.

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

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

5.3 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.

$I_H = (E_E) + (E_{E+H})$ (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.

$I = \sum (I_H)_k$, where k is n x m (5.12)

5.4 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.

$I_{-R-H} = (E_E)$ (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.

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

6 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.

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: (1) E_1 with E_2, R_{21} or R_{22}; (2) plus the match of R_{11} with E_2, R_{21} or R_{22}; or (3) 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.

The 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. 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.

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.

7 Conclusions

This work 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. 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 chose to use hierarchies, because they 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.

Using a simulation, 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.

This work 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.

8 References

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

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

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

Fonseca, F. (2001) *Ontology-Driven Geographic Information Systems*. Ph.D.Thesis, University of Maine.

Fonseca, F., Egenhofer, M., Davis, C., and Câmara, G. (2001) Semantic Granularity in Ontology-Driven Geographic Information Systems. *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity*: pp. to appear.

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

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

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

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

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

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

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): pp. 202-225.

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

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

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

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.

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

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

Sondheim, M., Gardels, K. and Buehler, K. (1999) GIS Interoperability. in: Longley, P., Goodchild, M., Maguire, D., and Rhind, D., (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): pp. 83-106.

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

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

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

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

Vckovski, A. (1998) International Journal of Geographical Information Science - Special Issue: Interoperability in GIS. 12(4): pp. 297-298.

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

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