USING ONTOLOGIES FOR GEOGRAPHIC INFORMATION INTEGRATION

Frederico Torres Fonseca
The Pennsylvania State University, USA

Keywords: ontologies, GIS, geographic information integration, interoperability

Contents

1. Introduction
2. Ontologies and Interoperability
2.1. GIS Interoperability
2.2. Ontology and Interoperation
2.3. Ontology Levels
3. A Conceptual Framework for Geographic Information Integration
3.1. Abstractions of the Geographic World
3.2. The Use of Multiple Ontologies
4. Ontology Integration
5. The Next Generation of Information Systems
Acknowledgments
Glossary
Bibliography
Biographical Sketch

Summary

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. Often 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.

Ontologies are used as the foundation for the integration of geographic information. By integrating ontologies that are linked to sources of geographic information, this information can be integrated based primarily on its meaning. The use of an ontology, translated into an active, information-system component, leads to ontology-driven information systems (ODIS) and, in the specific case of GIS, it leads to what is called ontology-driven geographic information systems (ODGIS).

The framework presented here is a foundation for future ODIS. The use of a highly semantic approach follows a trend of modern information systems. Although the

development of the large amount of ontologies necessary to fully unleash 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 (2002) keyword based systems.

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 article. This integration is usually 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 from new and sophisticated data collection technologies. Now information integration is turning into a science, and it is necessary to find innovative ways to make sense of the huge amount of information available today (see *Advanced Geographic Information Systems*).

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. The term semantics refers 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, the geographic world. A concept such as a 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. 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. Information systems based on ontologies use the second option, because their goal is to integrate the information that represents the view of diverse communities, each one with its own ontology. These different views, expressed as ontologies, can be integrated across different levels of detail.

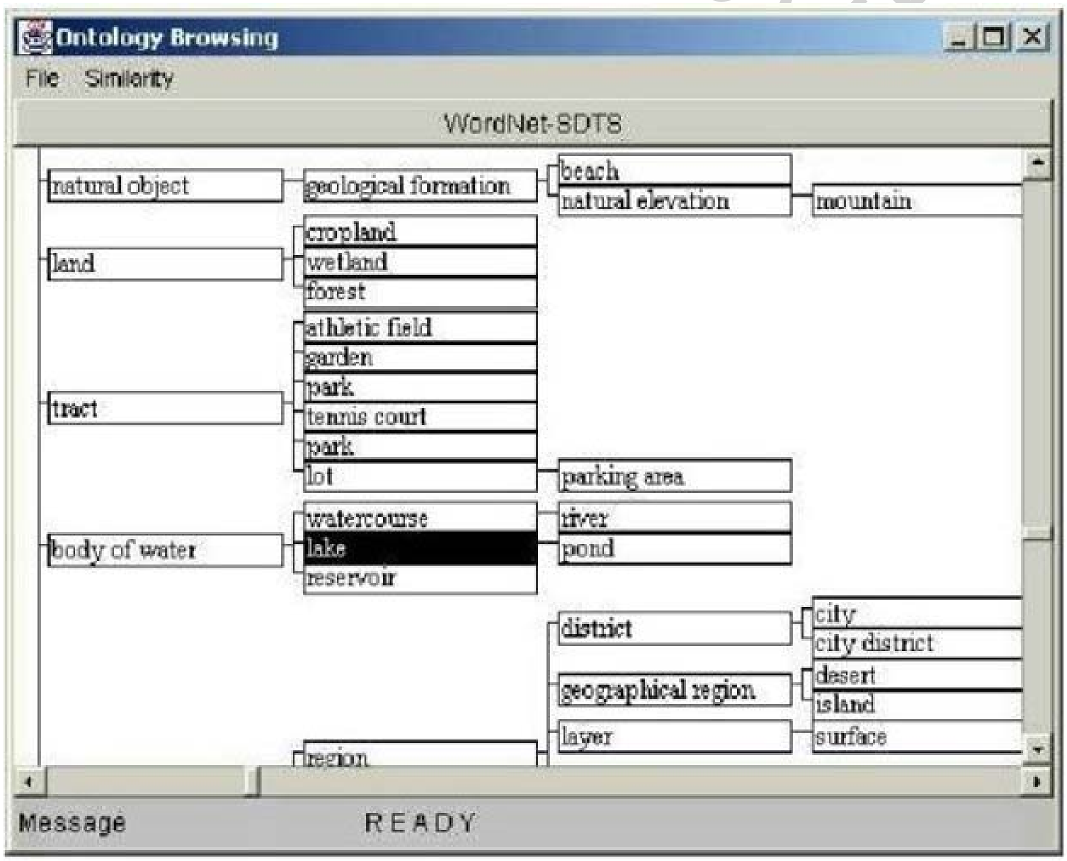

Figure 1. A subset of the combination of the WordNet and SDTS ontologies

For instance, a lake can be seen differently by distinct groups of people. For a water department a lake can be a source of pure water. For an environment scientist it is a wildlife habitat. For a tourism department it is a recreation point. A selection of the combination of the WordNet ontology and the ontology extracted from SDTS can be seen in Figure 1. From the point of view of this ontology a lake "is a body of water surrounded by land." This ontology can be considered as a high-level ontology. Therefore, in an ontology-based framework, different concepts of lake should be derived from this high-level ontology 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 users think are relevant to their concept of lake. The same happens with the other groups of users. If they all inherit from the WordNet–SDTS lake they will be able to share complete information at this level only, although they can share partial information at lower levels.

Ontologies are used as the foundation for the integration of geographic information. By integrating ontologies that are linked to sources of geographic information, geographic information can be integrated based primarily on its meaning. The use of an ontology, translated into an active information-system component, leads to ontology-driven information systems (ODIS) and, in the specific case of GIS, it leads to what is called ontology-driven geographic information systems (ODGIS). In this kind of system, the entities in the ontologies are translated into software components (for example, classes in the Java programming language).

2. Ontologies and Interoperability

The availability of network resources and the globalization of the economy lead to an increasing need to integrate and reuse information originating from diverse sources. Besides that, in the GIS community, the increasing availability of spatial information in the form of a huge amount of information gathered about the earth, not only from new spatial information systems, but also from new and more sophisticated data collection technologies, created the need to share geographic information. During the last few years, data from 1-m resolution satellites have become commercially available, and unmanned aerial vehicles can provide us with aerial video over rapidly-evolving focused scenes.

Heterogeneity research is not specific to the GIS realm, but 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 (that is, fields and objects) and different computer representation of these entities (that is, raster and vector). (See *Spatio-temporal Information Systems* and Section 3.2.) Sophisticated structures, such as ontologies, are good candidates for abstracting and modeling geographic information. The support and use of multiple ontologies is a basic feature of the modern information systems that address the semantic issue of information integration. Ontologies capture the semantics of information and they can be used to store the related metadata enabling in this way the semantic approach to information integration.

2.1. GIS Interoperability

Despite initiatives such as SDTS, SAIF, and OpenGIS (see *Geospatial Interoperability: the OGC Perspective*), the use of data transfer standards as the only worthwhile effort to achieve interoperability is not widely accepted (see *Spatial Data Standards*). Since widespread heterogeneity arises naturally from a free market of ideas and products, it is difficult for standards to banish heterogeneity by decree. The use of semantic translators in dynamic approaches is a more powerful solution for interoperability than the current approaches that promote standards (for a complementary discussion on interoperability see *GIS Interoperability, from Problems to Solutions*).

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:

- 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.

There is a wide range of problems when integrating different geographical databases. For instance, when a query is executed that must fetch data from multiple databases, it must be routed only to sites that are likely to contribute to the answer. The problem is that these sites might be fragmented geographically or thematically. Therefore, it is not only necessary to align data (geographic) boundaries, but it is also important to ensure their topological continuity. One of the solutions for this problem is the creation of local and global spatial indices. A river or a road can run across different states or even different countries. Such features must be linked to a central index that has the key to reconstruct them when necessary (see *Spatio-Temporal Information Systems*).

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. 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. 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. Other lines of research in interoperability consider different solutions such as the use of ontologies as the common point among diverse user communities. The use of ontologies to enable interoperation is the theme of the next section.

2.2. 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 project that uses geographic data and can also be a positive factor in the success of a project. 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 the user can rely on previous labor and experience. Many high-level government institutions recommend the use of mechanisms that enhance the possibility of information sharing.

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. Ontologies are crucial for knowledge interoperation, and they can serve as the embodiment of a consensus reached by a professional community. Sharing the same ontology is a precondition 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 who made the information available.

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 achieved through description logic. Another semantic approach is a similarity assessment among ontologies using a feature-matching process and semantic distance calculations. 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 ODIS? Users of geographic information can be grouped into geospatial information communities according to their conceptualizations of the world. The definition of a geospatial information community should not be restricted to users that share the same data model. Hence the definition of such a community as a group of users that share an ontology, and that can, furthermore, 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 geospatial information communities. Semantic translators, also called mediators, 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.

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 (for other kinds of solution see *Interaction Issues and Decision Support in Intelligent GIS, Web-based Spatial Decision Support: Technical Foundation and Applications*).

-
-
-

Bibliography

Bishr Y. (1997). *Semantic Aspects of Interoperable GIS*. Doctoral thesis, Wageningen Agricultural University, the Netherlands. [This thesis addresses the importance of semantics in the integration of geographic information.]

Egenhofer M., Glasgow J., Günther O., Herring J., and Peuquet D. (1999). Progress in computational methods for representing geographical concepts. *International Journal of Geographical Information Science* **13**(8): 775–796. [A review of the efforts towards the development of GIS science in the period 1990–2000. This paper focuses on the computational aspect of this problem.]

Fonseca F. (2001). *Ontology-Driven Geographic Information Systems*. Doctoral Thesis, University of Maine, Orono, USA. [This thesis gives the foundations of ontology-driven GIS.]

Goodchild M., Egenhofer M., Fegeas R., and Kottman C. (1999). *Interoperating Geographic Information Systems*, 536 pp. Norwell, MA: Kluwer Academic Publishers. [This book has an extensive report on the latest technologies used for the integration of GIS. It contains extensions of the works presented at Interop'97, the first international conference on Interoperating Geographic Information Systems.]

Gruber T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies* **43**(5/6): 907–928. [This paper gives an overview of the remarkable work of the Knowledge Systems Laboratory, Stanford, in the area of ontologies.]

Guarino N. (1998). *Formal Ontology in Information Systems*, 347 pp. Amsterdam, the Netherlands: IOS Press. [This book provides the most comprehensive reference for the use of ontologies in a variety of fields in computer science.]

Smith B. and Mark D. (1998). Ontology and geographic kinds. *International Symposium on Spatial Data Handling*, pp. 308-320. Vancouver, Canada: Taylor and Francis. [This paper describes the fundamentals of the ontology of geographic space.]

Wiederhold G. (ed.) (1996). *Intelligent Integration of Information*, 216 pp. Norwell, MA: Kluwer Academic Publishers. [This book show different alternatives for information integration in modern information systems.]

Winter S. (2001). Ontology in the geographic domain. *International Journal of Geographical Information Science: Special Issue* **15**(9) [This special issue presents different views of using ontologies in the geographic domain, and has extended versions of work presented at the Euro Conference on the Ontology and Epistemology for Spatial Data Standards, September 22–27, 2000, at La Londe-les-Maures, France.]

Vckovski A. (1998). Interoperability in GIS. *International Journal of Geographical Information Science: Special Issue* **12**(4). [This special issue gives an overview of the most recent approaches to GIS interoperability.]

Biographical Sketch

Frederico Torres Fonseca is an assistant professor at the School of Information Sciences and Technology at the Pennsylvania State University. He graduated from the Universidade Federal de Minas Gerais in 1977 with a degree in Computer Science. He also graduated from the Universidade Católica de Minas Gerais in 1978 with a degree in Mechanical Engineering. From 1977 to 1998 he worked as a system analyst in several companies in Brazil. For over 10 years he worked as a GIS analyst at PRODABEL—the Municipal Data Processing Center of the city of Belo Horizonte, Brazil, where he participated in a large pioneer GIS project in South America for municipal management. In 1997 he got his Master's degree in Public Administration and Computer Science at Fundação João Pinheiro, Brazil. Dr. Fonseca was a graduate research assistant at the National Center for Geographic Information and Analysis from 1998 to 2001. He was the recipient of the 1999 ESRI/IGIF Scholarship, a NASA/EPSCoR fellowship, and the 2000 Graduate Research Assistant Award of the College of Engineering of The University of Maine. He got his Ph.D. degree in Spatial Information Science and Engineering from the University of Maine (2001). Dr. Fonseca's main research interests are the development of interoperable GIS frameworks, ontology management, and spatial decision support systems.