Interoperability in Geospatial Information Systems

Hassan A. Karimi
University of Pittsburgh, USA

Ratchata Peachavanish
University of Pittsburgh, USA

INTRODUCTION

Geospatial information systems (GIS) are an important sector of the information industry, as well as an essential component of the information technology infrastructure (Lo & Yeung, 2002). They are a type of computerized information system specifically designed and used to solve *geospatial* problems, those which are related to locations on the surface of the earth (Longley, Goodchild, Maguire, & Rhind, 2001). The extent of usefulness of GIS has been proven across many diverse applications in many disciplines. They have long been used in traditional application settings, such as land management and natural resources, and have recently become an important element in emerging applications, for example, in ubiquitous mobile computing environments.

Since the mid-1990s, the focus of computing has shifted from stand-alone and locally networked environments to wide-scale, distributed, heterogeneous computing infrastructures. This coupled with the exponential growth in the use of the Internet has enabled and compelled new ways of using GIS. A wide range of GIS applications are now widely available on the Internet for users from anywhere in the world. In addition, the proliferation of wireless and mobile computing technologies, such as cellular phones and Personal Digital Assistants (PDA), has provided new platforms and paved the way for the emergence of new GIS applications. Because of the advances in computing, GIS applications are now designed, implemented, and applied very differently from their predecessors.

Contrary to their past monolithic design and implementation, GIS are now becoming an integral component of many diverse software packages specifically designed for solving problems in different application domains. In addition, current computing trends suggest that future GIS will be multitiered and used in heterogeneous network environments, where computers of different platforms coexist and tasks are performed in a distributed manner. Consequently, future GIS will have to be *interoperable*—they will have to be able to work together in a seamless fashion.

BACKGROUND

Efficient and effective use of GIS to solve geospatial problems generally requires special skills. Today's GIS platforms are mostly designed for workstations or personal computers and provide generic "toolbox" *geoprocessing* operations that can be broadly applied to many problems in different application domains. To utilize these packages, the users must possess certain knowledge and skills. First, they must have knowledge about how real-world objects are represented in GIS; for example, the boundary of a county is represented as a set of points that defines a polygon, and a railroad is represented as a line object. Second, the users must know the range of available geoprocessing operations in GIS and how they are applied to solve geospatial problems. For example, the users must know that in order to determine if a railroad crosses a county boundary, a geometric intersection operation using a polygon and a line that define the county boundary and the railroad, respectively, should be applied. In addition, knowledge about geospatial data sources, geospatial data storage, and methods of obtaining geospatial data are also needed. To solve the problem of county boundaries, the users must have knowledge about the sources of data sets for polygons representing county boundaries. They must also know the format and structure in which the data sets are stored. Lastly, the users must know how a GIS software package operates and how to use it to solve problems. For example, they must know how to operate the ArcInfo GIS software package and be familiar with the methodology in ArcInfo for incorporating data sets into the project, including format conversion, coordinate transformation, and importing procedures. They must also know the specific commands and syntax for invoking geoprocessing operations in ArcInfo as well as the specific behavior of each operation (e.g., Does the "intersect" operation provided by ArcInfo partially, or fully, solve the problem?).

These difficulties in using GIS are due to a number of historical and practical reasons. GIS software packages generally approach geospatial problems in terms of abstract geometrical objects and operations, that is, computations on points, lines, and polygons. This imposes a heavy burden on the users because the first task in solving any problem with GIS is to map real-world problems into an environment where GIS techniques and tools can be used. This task is further complicated by the fact that problems in different application domains are often treated differently in GIS. For example, applications in ecology usually involve large-scale raster data and spatiotemporal analysis for visualization purposes, while applications in urban navigation generally involve a smaller geographic extent and are concerned mainly with real-time decision information. In addition, GIS software packages have historically been developed independently with little regard to data sharing (Goodchild, Egenhofer, & Fegeas, 1998). Different GIS software packages use their own proprietary formats, schemas, and terminologies to represent geospatial data and concepts. This has exacerbated the issue of data use, especially when they are to be shared, requiring manual conversions or availability of import and export tools. This process is often nontrivial and, considering the large volumes of data commonly required in GIS projects, is also very time consuming.

MAIN THRUST OF THE ARTICLE

Information Heterogeneity

The aforementioned issues are related to interoperability. The basis for problems related to interoperability is *information heterogeneity*, which is divided into three levels (Sheth, 1999): *syntactic heterogeneity*, which refers to the differences in formats and data types; s*tructural heterogeneity*, which deals with the differences in data-modeling constructs and schemas; and *semantic heterogeneity*, which refers to the variations of the intended meanings of concepts and terminologies. Table 1 provides examples of information heterogeneity in GIS.

The issues of syntactic and structural heterogeneities have been extensively addressed in the past within the computer and information science discipline. Recently, much research has been focused on addressing the issue of semantic heterogeneity, which is a significant problem in the field of GIS. In general, semantic heterogeneity is a result of different conceptualizations and representations of things in the world and can be distinguished into two types (Bishr, 1998).

Cognitive heterogeneity, which arises when two groups of people from different disciplines conceptualize the same real-world facts differently. As an example, a geologist thinks of hill slopes as areas where soil erosion or landslides can occur, but a tourist manager may think of hill slopes as areas where skiing is possible (Dehn, Gartner, & Kikau, 1999).

Naming heterogeneity, which arises when different names are used for identical concepts of real-world facts. For example, *hill slope* is also known as *valley side*, *mountain flank*, or simply *slope*.

Due to the widespread use of GIS by users both within and across disciplines, semantic heterogeneity in GIS is increasingly becoming an important issue in the GIS community. In the first example illustrated by Lutz, Riedemann, and Probst (2003), the semantic of the *touch* topological operator in the GeoMedia Professional GIS software package is different from that of Oracle 9i Release 2 Spatial (Table 2.). In GeoMedia, two polygons would satisfy the touch operator if their boundaries and/or interiors intersect. In Oracle, on the other hand, two polygons would satisfy the touch operator only if their boundaries, and not their interiors, intersect.

Furthermore, two GIS software packages may use

Table 1. Information heterogeneity in GIS

Information Heterogeneity	**Examples**
Semantic	Different behaviors of the "intersect" operation from different GIS software packages Different interpretations of the word "within" in a user's query
Structural	Different data dictionaries when merging two or more data sets Different metadata standards
Syntactic	Different data formats (e.g., Shapefile, ASCII [American Standard Code for Information Interchange], XML [eXtensible Markup Language])