Contents

1. Introduction

1.1 The Necessity of Sharing Geodata

Sharing geodata is necessary for most of spatial or spatial-temporal analysis. Information from different sources are often required to effectively manage the task at hand and facilitate better reasoning and decision for territorial policy making and planning. Taking street management as an example, street repair works require information from different sources, such as, information from the local water company, gas company, electricity company, telephone company and municipal service in charge of sewerage. For environmental control geotechnical data (soil and sub-soil), meteorological data, hydrologic data, etc are required. All these examples illustrate the necessity of sharing geographic information.

1.2 The Importance of Harmonisation of Geodata

Owing to the necessity of extracting all essential and complete information for quality geo-referenced applications, it is important to have successfully integrated geoinformation. In order to avoid duplication and conflict of the existing data and benefit from the newest updates or information located in other databases, harmonization of geodata at different levels, namely, syntactic, schematic and semantic levels, is necessary. Another reason for harmonization geodata is the cost reduction due to the possibility of multiple uses of the data sources.

Let us take the latest generation of harmonization of geodata, which is semantic one, as an example. Imagine if a hydrologist wants to look for the information of water level of a certain river, she might possibly use 'water level' as the keyword for query, however, if the database she used adopts 'water depth' for the same piece of information, she would not be able to retrieve such information due to the different terminologies used by the requester and the provider. Hence, harmonization of geodata is needed so that all relevant information sources could be retrieved.

2. Overview of 3 Generations of Interoperability Concern of Information System

Interoperability has been a basic requirement and concern for information system environment for decades. How have key requirements and elements of interoperability changed over time? What are the chief progresses and achievements? What is left to be researched and improved? The following table (Sheth, 1998) summarizes basically 3 generations of interoperability concern of information systems.

	Generation I	Generation II	Generation III
Level of Interoperability Concern (emphasis underlined)	System, data	System, data, information	System, data, information and knowledge
Types of Interoperability emphasized	System (computer system and communication), limited aspects of syntax and structure (data model)	Syntax (data types and formats), structure (schematic, query languages and interfaces)	Semantic (increasingly domain-specific)
Scope of System Interoperability	Handful of interconnected computers and databases	Tens of systems on a LAN, databases and text repositories	Enterprise-wide, globally and mobile

	Generation I	Generation II	Generation III
Types of Data	Structured databases and files	Structured databases, text repositories and semi-structured data	All forms of digital media with increasing support for visual/spatio-temporal/scientific/engineering
Data/Info. Interoperability Approaches	Structural and data modal, data representation	Understanding of a variety of metadata, comprehensive understanding of schematic heterogeneity	Comprehensive use of metadata, increasing emphasis on semantics and ontology supported approaches
Interoperability Techniques (representative samples)	Data-level relationships, common data models, database exchanges and schema integration	Schematic and metadata-level relationships, wrappers, extractors, single ontology and multidatabase consistency, mediators	Multiple ontologies, information/semantic level relationships, context, media-independent information correlation, inter-ontological relationships, metadata consistency
Representative Applications	Integrations of business databases or public databases	Digital library, integrated access to heterogeneous data for a software team	Digital earth, environmental phenomena, multi-step and multi-modal intelligence analysis, navigation application

Table 1 An Overview of Three Generations of Interoperability of Information Systems

2.1 Generation I – System

The scope of interoperability of the information system during the first generation was primarily departmental and focused basically just within an enterprise. By the 1980s, corporations had huge amounts of data stored in different departments for different purposes and applications. They were stored on computers with different hardware and software (including DBMS). Because of the need for exchanging and sharing data between departments within corporations, ways for solving such heterogeneity is required. The representative work during this stage is the handling of systems of multidatabases, addressing particularly the heterogeneity due to the differences in DBMS, however, the scope was limited to just few databases or computer nodes in a local area network.

2.2 Generation II – Data (Syntax and Structure)

With significant impact of the internet and the Web in 1990s, the scope of interoperability during this generation had extended from within enterprises to enterprise-wide (between enterprises). It was common to have tens of computers and data repositories involved in this generation. The interoperability of systems at this stage looked not only at the structured data, but the semi-structured one, text and digital media. A broad variety of metadata is used in support of the integration.

2.3 Generation III – Information and Knowledge (Semantic)

With the advance of communication technology and global information structure, the dimension of distributing data/information is significantly extended, from a single system to global. More distribution and more autonomy of accessing information means at the same time, higher chance of the heterogeneity of the data. Key challenge in the third generation remains at semantic or contextsensitive level, where people expect the information systems helping them not just at the data level, but also the information level and increasing knowledge level.

3. Barriers to GIS Interoperability

In context of geographic information, people categorize, classify different real world objects and facts, give them names and create mental models (schema) and then present them in a database with the underlying syntax (Bisher, 1998). However the whole process gives rise to the following types of heterogeneity in the context of a database.

3.1 Syntactic Heterogeneity

Syntactical investigation, simply put, deals with the grammar of the schema and semantic expressions, without the necessity of knowledge of its meaning. Different grammatical structures used result in syntactic heterogeneity. In the context of GIS and database, each database may be implemented in a different DBMS of different paradigm, e.g. relational or object-oriented models. Different databases have different query languages (SQL or OQL). Syntactic heterogeneity also relates to the difference in the geometric representation of geographic objects, e.g. raster or vector representation.

3.2 Schematic Heterogeneity

If an object in one database is regarded as the property of an object in another, it results in the schematic heterogeneity or object classes have different generalization hierarchies, although in fact they represent the same real world facts, e.g. a street may be classified as a major street in a system due to its heavy traffic load, while the same street would be classified as a minor one in another system owing to its width.

3.3 Semantic Heterogeneity

People could possibly relate different expressions or vocabularies to the same real world features or facts, e.g. one might use different terminologies for ‘water level’ in a database, such as ‘water depth’ or ‘watermark’ and hence resulting in the semantic heterogeneity.

Semantic heterogeneity occurs at different levels. At the metadata level, semantic heterogeneity impedes the discovery of GI. At the schema level, semantic heterogeneity obstructs the retrieval of GI. At the data content level, semantic heterogeneity hinders the interpretation, integration and exchange of GI.

4. Recent Developments in Resolving Semantic Heterogeneity in GIS

4.1 Ontology-based Retrieval of Geographic Information

There are different ontology approaches to domains of GI for semantic organization and discovery, namely, multiple ontology approaches; single ontology approaches and hybrid approaches. In multiple ontology approaches, each information source has its own local ontology, several local ontologies do not share the same vocabulary, which makes it difficult to compare different application ontologies. In single ontology approaches, the shared vocabulary in a single catalogue is introduced in order to specify the semantics of all information sources and queries. These semantic descriptions sharing similar view on a domain would be shared by all requesters. Like single ontology approaches, hybrid approaches use also a global shared vocabulary. The shared vocabulary in this approach consists of basic terms of a domain. They can be combined to describe the semantics of each information source or query in separate application ontologies. The shared vocabulary can have different levels of abstraction of domain, e.g. domain of ‘Measurement’ would be more abstract than that of ‘Hydrology’, however, both domains can be stored in the shared vocabulary.

For both single ontology and hybrid approaches, the semantics of the shared vocabulary are assumed to be understood by both the providers as well as the requesters, hence no further definition should be required.

M. Lutz et al. (2009) use hybrid ontology approaches for their application as it is considered flexible in building application ontologies, as in multiple ontology approaches. Meanwhile, they remain comparable, as in single ontology approaches.

4.2 Mediator and Wrapper Technologies

Technically, mediators and translators (wrappers) are middleware (software and programs) distributed over a network, which enable the resolution of semantic conflicts between data providers (information sources) and data receivers (users). Using local ontology and a reasoning system, the combination of these technologies enables detecting and correcting the discrepancy in semantics between queries and information sources. The type of semantic conflict can be identified and reformulated, then the requested data can be located in the information sources and results are returned to the users.

5. Examples and Applications

Hübner et al. (2004) illustrated a bike trip example, which shows one of the ontology-based applications developed by GeoShare. GeoShare uses the information broker middleware called BUSTER (Bremen University Semantic Translator for Enhanced Retrieval), which was developed at the Center for Computing Technologies (TZI) in Bremen. It supports intelligent search and semantic data integration in distributed environments.

In this bike trip example, a tourist intends to cycle along a certain river in Bremen, Germany and at the same time, tries to locate a nice place for fishing and bathing. Hence he wants to know the water quality of the river. BUSTER allows conceptual semantics search, spatial and temporal search, however, the temporal issue is not the concern in this example. With BUSTER middleware, domain, ontology and spatial data model (in this case – Germany’s NUTS model) could be selected. All terminologically relevant information (all vocabularies defined in domain ‘Hydrosphere’) for the requested geographic location (in this case – Bremen) can be retrieved. When the tourist selects all resulted vocabularies, the corresponding information sources (e.g. Environmental Department at the Senate of Bremen) would be retrieved. With a mapserver, the geodata can be represented on an interactive map.

Another example of harmonizing and integrating geodata was illustrated by Volz (2007). For navigation applications, it is also important to have consistent spatial representations of the same real world objects in different spatial databases, nonetheless, this is hardly the case in reality. He proposed using the information stored in MRep (Multi-Representations) Relations for analyzing the objects of multiple spatial representations, according to their explicit relations, such as geometric; topological and thematic similarity. The semi-automated software, called Relation Builder Toolbox had been used for such purpose. The shortest-path algorithm was developed for searching the shortest path in multi-representation databases. The result showed that if the application requires high reliability of the resulting way, the shortest path would end up actually longer than that for the application, where a certain degree of uncertainty of the resulting path is allowed.

Hossain and Rashed (2007) presented an application of a proposed architecture solving the semantic heterogeneity problem in different data sources. The application concerns the reconstruction of the sewerage lines in Chittagong city in Bangladesh. This might affect also the telephone lines, pipe lines and power supply lines. In order not to put its inhabitants’ lives at risk, before the reconstruction, information access to these relevant databases is necessary. However, due to the lack of coordination between these authorities, it is hard to expect all these information systems use the same terminologies as those in the system (WASA) of the water and sewerage authority, hence the problem of semantic heterogeneity presents when these systems are integrated. Hossain and Rashed (2007)'s approach was to use an ontological architecture with global community (GC), which contains an integrator, global schema and common ontology/mediator for resolving semantic conflicts instead of using a reasoning system and a mediator. By using an ontology editor, called OilEd 3.5, developed by the University of Manchester in 2004, a global schema definition is formed by taking into account of the local schema definition of each GIS of each concerned authority. When a request from WASA is received, the global schema of GC will look for the similar attribute and pass this information to the common ontology/mediator, which will use some rules to retrieve the relevant data from the associated local schemas (telephone; gas and power supply). The retrieved information will then be translated in GC in a format that could be interpreted by the initial requester (in this case: WASA), hence the system can then possibly use the acquired information for further action.

6. Conclusions and Outlook

Sharing of knowledge, i.e. geodata is one of the crucial issues in GIS. However, different people very often hold different views or interpretations of the same feature/data in the real world we are living in. The reason for this could be due to their personal experiences/common sense or that they use the concepts or terminologies in their respective field of expertise. Because of these differences the problem of semantic heterogeneity during the data sharing process occurs. Data harmonization is hence necessary for an efficient and effective retrieval of geo-information.

In this context, integrating geo-information from different sources for such purpose is, however, never an easy task. Problems occur at different levels during the data integration process. Efforts in improving the information systems in such regard have been made and have evolved over decades, from Generation I, dealing with only the systems to the latest semantic one. For the latest semantic issue, different approaches have been attempted, say, multiple ontology-based; single ontology-based; and the latest one concerning hybrid ontology-based retrieval of geographical information, which has been introduced by M. Lutz et al. in 2009. However, for such implementation, from the data providers' point of view, certain complexity of creating and registering application ontologies still remains. Future work should consider how the process of creating formal descriptions of the geodata can be automated.

References:

Bishr, Y. (1998). "Overcoming the semantic and other barriers to GIS interoperability." International Journal of Geographic Information Science 12(4): 299-314.

Egenhofer, M. J. (2002). "Toward the semantic geospatial web." In: The 10th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS). McLean, VA, USA: 1-4.

Fonseca, F. T., M. J. Egenhofer, et al. (2000). "Ontologies and knowledge sharing in urban GIS." Computers, Environment and Urban Systems 24(3): 251-272.

Hakimpour, F. and S. Timpf (2001). "Using Ontologies for Resolution of Semantic Heterogeneity in GIS." In: The 4th AGILE Conference on Geographic Information Science. Brno, Czech Republic: 385-395.

Hübner, S., R. Spittel, et al. (2004). "Ontology-based search for interactive digital maps." IEEE Intelligent System 19(3): 80-86.

Lutz, M. and E. Klein (2006). "Ontology-based retrieval of geographic information." International Journal of Geographic Information Science 20(3): 233-260.

Lutz, M., J. Sprado, et al. (2009). "Overcoming semantic heterogeneity in spatial data infrastructures." Computers and Geosciences 35: 739-752.

Hossain, M. S. and Rashed M. (2007). “Resolving geo-spatial semantic conflicts – an interoperability issue.” Humanomics 23(2): 102-109.

Sheth, A. P. (1998). “Changing focus on interoperability in information systems: from system, syntax, structure to semantics.” in: M. F. Goodchild, M. J. Egenhofer, R. Fegeas and C. A. Kottman, eds, Interoperating Geographic Information Systems, 1998: 5-30

Vögele, T. and R. Spittel (2004). "Enhancing spatial data infrastructures with semantic web technologies." In: The 7th Conference on Geographic Information Science. Heraklion, Greece: 105-111.

Volz, S. (2007). "Shortest Path Search in Multi-Representation Street Databases." in: Georg Gartner, William Cartwright and Michael P. Peterson, eds, Location Based Services and TeleCartography, 2007: 165-178