Proceedings of the 6th AGILE
Dates – Lieu

REALIZATION OF GIS SEMANTIC INTEROPERABILITY IN LOCAL COMMUNITY ENVIRONMENT

Leonid Stoimenov, Slobodanka Djordjević-Kajan

CG&GIS Lab, Department of Computer Science, Faculty of Electronic Engineering, University of Niš, Yugoslavia

1. INTRODUCTION

Popularity of GIS in government and municipality institutions induce increasing amount of available information [1]. In local community environment (city services, local offices, local telecom, public utilities, water and power supply services, etc) different information systems deal with huge amount of available information, where the most of data in databases is georeferenced. Information that exists in different spatial databases may be useful for many other GIS applications. But, information communities find it difficult to locate and retrieve data from other sources, in reliable and acceptable form. In such systems reuse for geodata are very often a difficult process, because of poor documentation, obscure semantics, diversity of data sets, and the heterogeneity of existing systems in terms of data modelling concepts, data encoding techniques and storage structures [2]. Also, available information is always distributed and no one wants to share his own information in public without commitment. In that case, centralized control is not applicable and not practical, since the ownership of data is in domain of organizations whom they belong, and no one wants to share his own information with public. Because of that, there is a need to provide communication and collaboration between these information sources without centralized control. The problem of bringing together heterogeneous and distributed information systems is known as interoperability problem.

The realization of interoperable GI systems is weighty process, as a consequence of two main system characteristic - distributed data sources and their heterogeneity [3]. Information systems heterogeneity may be considered as structural (schematic heterogeneity), semantic (data heterogeneity), and syntactic heterogeneity (database heterogeneity) [4]. Syntactic heterogeneity means that various database systems use different query languages (SQL, OQL, etc). Structural heterogeneity means that different information systems store their data in different structures. Semantic heterogeneity considers the content of an information item and its meaning. Semantic conflicts among information systems occur whenever information systems do not use the same interpretation of the information. Semantic heterogeneity of the data sources causes serious problems. Domain experts use the concepts and terminology specific for their respective field of expertise, and use different parameters and different languages to express their model of a concept. Humans use their "common sense", i.e. their knowledge about the world, to translate the meaning of foreign set of concepts and terms in their own terminology. Software systems usually do not have any knowledge about the world and have to explicitly be told how to translate one term into another.

The paper is structured as follows. In the second part, we describe related work in GIS interoperability and ontologies. The goals of our research activities, described in the third part of this paper, are defining architecture for semantic integration of distributed and heterogeneous GIS data sources and adding the integration technology to the existing framework. Making local geographic datasets available publicly and establishing a common interoperability framework over shared data interchange protocols are important parts of this research.

2. SEMANTIC INTEROPERABILITY AND GIS

The universal interchange of geospatial data has become one of the actual GIS topics to date. An increasing number of geodata producers and users in local community environment have expressed the need for interoperable GISs and for the integration of geodata. Each of these user groups has a different view of the world. From the perspective of geography these groups is named Geospatial Information Communities (GIC) [5]. Each GIC is a group of users that shares a digital geographic information language and spatial feature definitions, and people inside each community agree on the most basic concepts.

Today, research on interoperability solutions is the way to migrate away from the monolithic systems that dominate the GIS market [6]. The need to share geographic information is well documented [7]. Recent reviews of GIS interoperability and integration efforts can be found in [8,9].

One important initiative to achieve GIS interoperability is the OpenGIS Consortium. This is an association looking to define a set of requirements, standards, and specifications that will support GIS interoperability. The objective is technology that will enable an application developer to use any geodata and any geoprocessing function or process available on 'the net' within a single environment and a single workflow [5]. But, data standardization is not the whole solution. The interoperability problem would go away if every system always uses the same data model to represent the same information (identical names, structure, and representations).

Mediator-based system is important for spatial data interoperability architecture [10]. Mediator-based systems are constructed from a large number of relatively autonomous sources of data and services, communicating with each other over a standard protocol and enabling "on-demand" information integration [11]. Structural and syntactic heterogeneity may be solved by mediation. The 3-level architecture of mediator-based systems is constructed from an application layer, and large number of information sources (heterogeneous data sources with wrappers), communicating with each other over a standard protocol [11]. Nowadays, mediation concept is a part of the ARPA I3 (Intelligent Information Integration) reference architecture [12]. The I3 reference architecture should be seen as a vision of how vast amount of heterogeneous information can be incrementally pulled into a gigantic, reusable library of information resources.

The use of ontologies as semantic translators is possible approach to overcome the problem of semantic heterogeneity [13,14,15,16]. An ontology consists of logical axioms that convey the meaning of terms for a particular community. Logical axioms are the means to introduce concepts and their relations also express constraints on both concepts and relations. An ontology exists under a consensus by members of a community [14], e.g., users of single information system or people in one discipline. Recently, the use of ontology in information systems is discussed in [17] and specifically in GIS building in [2,9], and creation of GIS software components from ontologies in [18]. Research on ontology is becoming increasingly widespread in the computer science community, and its importance is being recognized in a multiplicity of research fields and application areas, including knowledge engineering, database design and integration, information retrieval and extraction.

3. GEONIS FRAMEWORK

In this section, we present the GeoNis semantic GIS interoperability framework based on mediators, wrappers and ontologies. The goals of research activities in GeoNis project are:

- defining interoperability architecture for integration of distributed and heterogeneous GIS data sources in local community environment,
- defining a methodology and software support for resolving semantic conflicts in data from different information sources,

GeoNis is framework for interoperability of GIS applications that have to provide infrastructure for data interchange in the local community environment. Data sources are local services and offices that own geodata in some format. Semantic interoperability in GeoNis is the ability of sharing geospatial information at the application level, without knowing or, understanding terminology of other systems. The problem of semantic heterogeneity in GeoNis will be resolved by concept of mediation and ontology. A semantic translation in GeoNis is developed for a particular domain, in our case for GIS applications in local city services, which deal with network data structures (local Telecom, water and soil pipe services, power supply services, and some local government services).

The total number of geodata providers in local community environment is indeterminable and unlimited. This implies the need for a flexible approach that can deal with the existing and the future geodata providers in interoperable systems. A standard model for spatial data is the first step to approach the solution for schematic and syntactic heterogeneity. The Open GIS Consortium (OGC) [5] specification aims to solve the problem of heterogeneity at the spatial data modelling level. Because of that, GeoNis uses OpenGIS standard as common data model to represent geodata on mediator level. Data models of local information sources are translated in common model using wrappers.

The goal of the GeoNis framework is to make the use of different data sources in their GIS applications simple for users. The following six presumptions we defined (modified from [19]):

- simple - users should not have to understand all details about the data or their source system to import and use them,
- transparent - complexities associated with data transfer should be hidden for users,
- open - interoperability should apply to all systems, and data exchange should be independent of the technology used,
- equal – systems are equal and autonomous,
- independence – systems have exclusive right to control its information and information processing without centralized control,
- effective - data transfer should be reliable, and the resultant data should be useful for the intended purposes,
- universal - all geospatial databases should be accessible,
- belonging – each system belongs to one GI community, and has its own institution, policy, culture and value viewpoint.

3.1 GeoNis Architecture

GeoNis is based on ORHIDEA [1,10] data integration platform. ORHIDEA platform has been developed in order to perform intelligent integration of information from multiple heterogeneous GIS (spatial and geographic), and non-spatial (thematic) data sources. ORHIDEA is a middleware, mediator system that provides data interchange and access to data sources, distributed over the Internet, without changing how or where data is stored.

The basic architecture of GeoNis framework is shown on Fig. 1. Each GIC (i.e. local service or office) contain GIS application and corresponding (spatial) database. For each data source there is a translator (or wrapper), which logically converts basic data objects to common information model. Translator/wrapper on destination GIC application converts request to local application data model (SQL or API). Data source application may execute these requests. The next layer performs mediator functions, which include transformation of data and mapping between data models. In order to make this logical translation, mediator converts queries to requests through information from common model and top-level ontology (TL ontology).

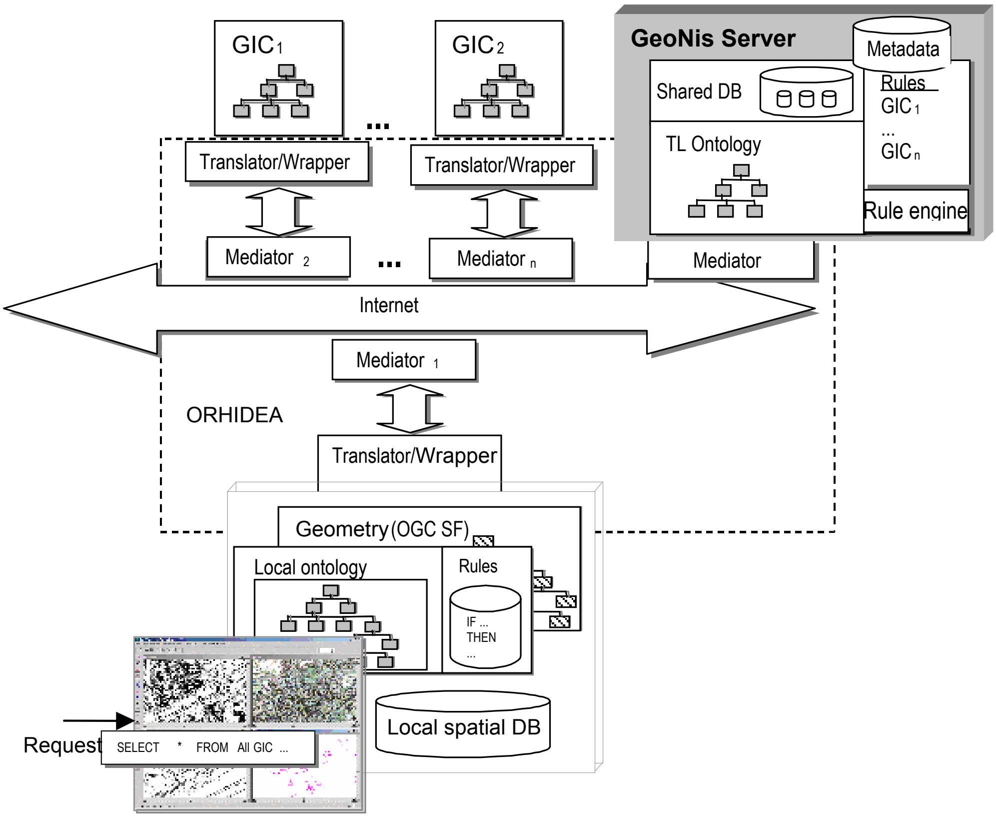

Fig. 1 Fig 1. GeoNis architecture

In addition to domain oriented GIS applications, there is one common GIS server that maintains all shared/common geographic data. Those data are public available and could be used by GIS clients in every GIC or citizens through the available public services on user demand. Data in local spatial databases are accessible in dependency of user privileges. Requests for specific data set are forward through local mediators or GeoNis server. GeoNis server also contains information about registered GIC and their access rights. Every new GIC who wants to participate in exchanging data must register on GeoNis server in order to allow access to his public available data and local ontology. After that, registered GIC have access to all available data from other public GIC databases (with possible given rights for access), and access to shared data on GeoNis server. GeoNis uses shared ontology, one for each source of data, i.e. for each GIC, and single top-level ontology located on GeoNis server. Local ontology is referring only to public available data.

3.2 Resolving semantic conflicts in GeoNis

GeoNis solution to the problem of semantic heterogeneity is to formally specify the meaning of the terminology of each GIC using local ontology and to define a translation between each GIC terminologies (local ontologies) and an intermediate terminology (in top-level ontology). GeoNis formal ontology consists of definitions of terms, and it includes concepts with associated attributes, relationships and constraints defined between the concepts and entities that are instances of concepts. In our system architecture it is assumed that the ontology is shared, and there exists commitment by the clients about data, which will be shared. But, in first phase, no need for commitment to common, top-level ontology. Intent of our formal ontology is for sharing, merging, and querying data, but not for reading and efficient processing.

In GeoNis we consider ontologies with inheritance relations *isa* and typed roles between concepts. An ontology is a triple $\mathcal{O}$=*(C, R, isa)* defined as follows:

1. $C=\{c_i \mid i=1,n\}$ is a set of concepts, where each concept c_i refers to a set of real world objects (concept instances with geo-representation),

2. $R=\{r_i \mid i=1,n\}$, is a set of binary typed roles (or relations) between concepts, defined as $R = \{(c_1, c_2) \mid c_1, c_1 \in C\}$

3. isa is a set of inheritance relationships defined between concepts, defined as $isa(c_1,c_2)=\{(c_1,c_2) \mid c_1,c_2 \in C \wedge E(c_1) \subset E(c_2)\}$

Inheritance relationships define a partial order over concepts and carry subset semantics. Semantic relationship between concepts c_1 and c_2 is based on their extensions. These extensions $E(c_i)$ of a concept c_i are defined as the set of real world objects, represented by concept c_i. If $E(c_1)$ is an extension of c_1 and is c_2 is a *super concept* of c_1, then $E(c_2)$ is an extension of c_2, i.e. all resources that are described by a concept description $E(c_1)$ are also described by a concept descriptor $E(c_2)$, where c_1 is a *sub concept* of c_2 (inclusion semantics of *isa*).

Ontologies can be represented as directed graphs where nodes correspond to concepts and arcs correspond to roles and *isa* relationships. A local (domain) ontology consists of definitions of terms from local terminology, organized in hierarchy or taxonomy of concepts. Concepts are terms from specific domain that is related to set or class of domain entities. Concepts are of two types: *primitive* and *composite* (or *non-primitives*). By primitive definitions, one expresses necessary constraints to be satisfied for instances in its extension. Non-primitive definitions are described by necessary and sufficient conditions. Non-primitive definitions can be used when one can give a thorough clear definition of a concept. Also, concepts could be *abstract* and *concrete*. Abstract concepts have description, but do not have geo-representation. Concrete concepts are real world objects (or entities) and they have geographic representation.

We have defined set of relations between concepts in ontology: (1) synonym, (2) hypernym, (3) hyponym, and (4) set of “topological” relations. Topology defines the spatial relationships between geographic feathures. In ORHIDEA ontologies, we define ”topological” relathionship between concepts which represents topological relations between real world entities. We’ve defined next ”topological” relathionship: T={arc-node, route, node-route, point-event}, where “arc-node” relationship defines that the line features can share endpoints (for example, concept “cable” with polyline representation and concept “cable equipment” with point representation), “route” means that line features can share segments with other line features, etc.

ORHIDEA follows the hybrid ontology approach, which means that a local ontology is constructed for each information source. The global terminology, represented by top-level ontology, can be seen as a set of basic terms of a domain. The relationship between concepts of different information sources is the task of the semantic inter-correspondences. We divide the semantic conflict (semantic inter-correspondences) into four types: semantic equality, semantic dissimilarity, semantic intersection and semantic contain.

Semantic equality (similarity) $SEqu(c_1,c_2)$ – means there is 1:1 map between description of concepts c_1 from ontology O_1, and concept c_2 from ontology O_2, and defined as $SEqu(c_1,c_2) = \{(c_1,c_2) \mid c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1)=E(c_2)\}$. This kind of relation is commutative and transitive.

Semantic dissimilarity $SNEqu(c_1,c_2)$- means there is no map between description of concepts c_1 (with name $Name(c_1)$) from ontology O_1, and concept c_2 (with name $Name(c_2)$) from ontology O_2, and $Name(c_1) = Name(c_2)$ (semantic dissimilarity between concepts with the same name). This kind of semantic inter-correspondence is important only for concepts with the same name and different domain’s semantic description, and defined as

$SNEqu(c_1,c_2) = \{(c_1,c_2) \mid c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) \neq E(c_2) \wedge Name(c_1)=Name(c_2)\}$. This kind of relation is commutative, but not transitive.

Semantic intersection $SIntersec(c_1,c_2)$- means there is 1:1 map between some part values in concept c_1 from O_1's domain and some part values in concept c_2 from O_2's domain (the sets of real-world objects represented by the concepts c_1 and c_2 overlap partially), and defined as $SIntersec(c_1,c_2)= \{(c_1,c_2) \mid c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) \cap E(c_2) \wedge E(c_1) \not\subset E(c_2) \wedge E(c_2) \not\subset E(c_1)\}$. This kind of relation is commutative, but not transitive.

Semantic contain $SContain(c_1,c_2)$– means for concept c_2 from O_2, every value in its domain has 1:1 map to the value in concept c_1 from O_1's domain, but not vice versa, and defined as $SContain(c_1,c_2)= \{(c_1,c_2) \mid c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) \not\subset E(c_2) \wedge E(c_2) \subset E(c_1)\}$. This kind of relation is neither commutative nor transitive.

Also, we define relationships between reference (common) model object classes and application ontology classes (concepts). The basic semantic relationship, abbreviated as *Refers_to(RefClass a, OntologyConcept c)*, is between concepts from local ontology and real-world classes (represented in common model). This kind of relationship defines the semantic of a geographic data set. With the *Refers_to* relationship we can define relationship between concepts from different application (local) ontologies, without existence of any other kind of semantic relationships.

Predicate *Refers_to* enables definition of *semantic relevance* inter-correspondence between concepts from different ontologies. Assume two geodata sets B and C, with ontology concepts b_1 and c_1 respectively, and reference model A, with class ***a***. There is a relation Semantic relevance $SRelev(b_1,c_1)$ between concepts c_1 and b_1 if there exist a class ***a***, such that class refers to both concepts c_1 and b_1, defined as: $SRelev(b_1,c_1)=\{(b_1,c_1) \in BxC \mid \exists a \in A \wedge Refers\_to(a,b_1) \wedge Refers\_to(a,c_1)\}$. With the *Refers_to* relationship we can define relationship between concepts from different application (local) ontologies, without existence of any other kind of semantic relationships.

The process flow of the data interoperability and semantic conflict resolution in GeoNis, described in this paper, is:

- user (data receiver) sends the query Q through translator,
- local translator send request to the mediator,
- using local ontology, the mediator judges and detects whether there is the semantic conflict between the receiver and the provider through semantic recognition. If there is the semantic conflict, the mediator identifies the type of semantic conflict,
- if there is the conflict, the mediator resolve the semantic conflict through semantic process, reformulate and get the right query Q',
- then, the mediator decides where requested data are located, and resend query (queries) to adequate information source(s) (data provider),
- data provider translator accesses the right information from the information provider, converts data to common model and returns the result to the local mediator,
- local (data provider) mediator sends data to the right mediator (data receiver mediator), which reformulates terminology in data using semantic inter-correspondences, and sends it to the user (data receiver).

4. CONCLUSION

We present here an ongoing case study and development of framework for semantic interoperability of GIS applications. This framework is aimed to resolve interoperability problem in local, municipality environment. The principles behind the ontology/mediation framework described in this paper are extensibility, relative autonomy of infrastructure nodes, and universal access to heterogeneous data sources from a variety of portals.

GeoNis project provides methodology and software support for the ontology mappings and resolving of semantic mismatches between terminologies according to the current context. Our solution is based on single, mediator-based architecture for interoperability in local community environment, OpenGIS Simple Feature as common model, and local ontologies for resolving semantic heterogeneity of data sources.

The specific contributions of our research are as follows:

- A generalized framework of an interoperable GIS environment is presented, in which both schematic and syntactic heterogeneity are resolved (by mediation and OGC standard as common model):
    - GeoNis uses widely accepted OpenGIS standard for geodata modeling and representation on mediator level.
    - Legacy data sources may be included in interoperability process; the only condition is realization of translator.
    - Changes in OpenGIS standard affects only to translators, not to information sources.
- Also, semantic conflicts can be detected and resolved. This framework is comprehensive enough to manage various types of semantic conflicts in heterogeneous information sources while preserving the autonomy of individual sources.

In the future, we will focus on research and technical problems of spatial mediation on the global scale, realization of methodology and tools for ontology development, research on semi-automatic recognition of semantic inter-correspondences between domain ontologies, developing of domain-oriented ontologies, and implementation of translators.

Acknowledgements

This research was partially supported by the project “Geographic Information System for Local Authorities based on Internet/WWW Technologies”, funded by Ministry of Science, Technology and Development, Republic of Serbia, Contract No. IT.1.23.0249A.

5. REFERENCES

[1] Stoimenov L., Djordjevic-Kajan S., Stojanovic D., «Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology», *Proc. MELECON 2000, 10th Mediterranean Electrotechnical Conf.*, Cyprus, Proc. Vol. 1, pp. 334-336, 2000.

[2] Devogele T., Parent C., Spaccapietra S., «On Spatial Database Integration», *International Journal of Geographic Information Science*, Vol. 12, No. 4, pp.335-352, 1998.

[3] Genesereth M., et al., «Infomaster: An Information Integration System», *ASIGMOD Conference*, pp.539-542, 1997.

[4] Bishr Y.A., «Overcoming the Semantic and Other Barriers to GIS Interoperability», *International Journal of Geographic Information Science*, Vol. 12, No. 4, pp.299-314, 1998.

[5] McKee L., and Buehler R., *The Open GIS Guide*, Third Edition, OpenGIS Consortium, Inc, 1998, http://www.OpenGIS.org/techno/specs.htm

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

[7] Vckovsky A. (ed.), *International Journal of Geographic Information Science - Special Issue: Interoperability in GIS*, Vol 12, No 4, 1998.

[8] Abel D.J., et all, «Towards Integrated Geographical Information Processing», *International Journal of Geographic Information Science*, Vol. 12, No. 4, pp.353-371, 1998.

[9] Laurini R., et al., «Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability», *International Journal of Geographic Information Science*, Vol. 12, No. 4, pp.373-402, 1998.

[10] Stoimenov L., Djordjevic-Kajan S., Milosavljevic A., «Realization of Infrastructure for GIS Interoperability», *YUINFO ‘02*, Kopaonik, Yugoslavia, 2002. (in serbian).

[11] Wiederhold, G. «Weaving Data into Information», *Database Programming & Design,* Vol. 11, No. 9, pp.22-29, 1998.

[12] *ARPA I3 - Reference Architecture for the Intelligent Integration of Information*, ARPA, Prepared by the Program on Intelligence Integration of Information (I3), version 1.0.1, May 19, 1995.

[13] Frank A.U., «Spatial Ontology: A Geographical Point of View», in *Spatial and Temporal Reasoning, O. Stock*, Ed. Dordrecht, Kluwer-Academic Publisher, pp. 135-153, 1997.

[14] Bishr Y.A., Pundt H., Kuhn W., and Rdwan M., «Probing the Concepts of Information Communities – A First Step Towards Semantic Interoperability», in M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman (eds.), *Interoperating Geographic Information Systems*, Kluwer Academic Publishers: Norwell, MA, pp. 55-70, 1999.

[15] Hakimpour F., Timpf S., «Using Ontologies for Resolution of Semantic Heterogeneity in GIS», *4th AGILE Conference on Geographic Information Science*, Brno, April 19-21, 2001

[16] Fonseca F., Egenhofer M., Agouris P., Camara G., «Using Ontologies for Integrated Geographic Information Systems», *Transaction in GIS*, vol. 6, No. 3, 2002.

[17] Guarino N., Ed., *Formal Ontology in Information Systems*. IOS Press, Amsterdam, 1998.

[18] Fonseca F., Egenhofer M., «Ontology-Driven Information Systems», *7th ACM Symposium on Advances in GIS*, Medeiros C.B. (ed.), Kansas Sity, MO, 1999, pp.14-19

[19] Levinsohn A., «Geospatial Interoperability: The Holy Grail of GIS», *GeoEurope*, available online: http://www.geoplace.com/gw/2000/1000/1000data.asp