FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. 17 (2002), 107–125

FRAMEWORK FOR SEMANTIC GIS INTEROPERABILITY

Leonid Stoimenov and Slobodanka Đorđević–Kajan

Abstract. This paper presents research in Geographic Information Systems (GIS) interoperability. Also, paper describes our work in development, introduces interoperability framework called *GeoNis*, which uses new technologies, proposed in this paper, to perform integration task between GIS applications and legacy data sources over the Internet. Our approach provides integration of distributed GIS data sources and legacy information systems in local community environment. The proposed framework uses the ORHIDEA mediation technology to allow communications between GIS applications over the Internet/Intranet. The problem of semantic heterogeneity will be resolved by concept of mediation and ontology.

1. Introduction

In recent years, a large number of diverse, distributed and heterogeneous information sources (databases, knowledge bases, collections of documents, etc), are available over the Internet. The exchange of information has become a crucial factor in today's economy. Many activities in business world involve different organizations that have to work together, and use existing information whenever possible, in order to reach a common goal. Similar situation is also in GIS and their applications.

Popularity of GIS in government and municipality institutions induces an increasing amount of available information ([28]). 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 are georeferenced. Information that exists in different spatial databases may be useful for many other GIS applications. But, information communities find

Received November 5, 2002.
2000 *Mathematics Subject Classification*. Primary 68U35.

it difficult to locate and retrieve data from other sources, in reliable and acceptable form. In such systems, geodata reuse is very often 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 ([10]). 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 ([3]). Because of that, there is a need to provide communication and collaboration between these information sources without centralized control. An effective management of cross–community technology development and integration requires new approaches ([5]). Also, changes to the existing GIS infrastructure are necessary. The problem of bringing together heterogeneous and distributed information systems is known as interoperability problem.

The realization of interoperable systems is weighty process, as a consequence of two main system characteristic – distributed data sources and their heterogeneity ([13]). Information systems heterogeneity may be considered as syntactic heterogeneity (database heterogeneity), structural (schematic heterogeneity), and semantic (data heterogeneity) ([6]). 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 interoperability is a very hard field, especially in geographical information system. Interoperability in general and semantic interoperability will lead to dramatic organizational changes in GI community. The advantages of successful information integration are obvious for many reasons:

– Quality improvement of data due to the availability of large and complete data.

– Improvement of existing analysis and application of the new analysis.

– Cost reduction resulting from the multiple use of existing information sources.

– Avoidance of redundant data and conflicts that can arise from redundancy.

The paper is structured as follows. In the second part, we describe related work in interoperability and mediation. The goal is to explore how these approaches can be scaled to the global interoperability context. The goals of our research activities, described in the third part of this paper, are defining architecture for integration of distributed and heterogeneous GIS data sources and adding the integration technology to the existing framework. We examine a research whose final goal is to make disparate data sources work together. The proposed platform uses agent–wrapper and mediator technology to allow communications between GIS applications over the Internet/Intranet. The problem of semantic heterogeneity will be resolved by concept of mediation and ontology.

2. Related Work

Research in information systems interoperability is motivated by the ever–increasing heterogeneity of the computer world. Heterogeneity in GIS is not an exception, but the complexity and richness of geographic data and the difficulty of their representation raise specific issues for GIS interoperability. Besides interoperability has to overcome complexity of sharing and integrating data between systems with different data structures and models, it also has to deal with semantic heterogeneity. This has become more important due to the fact that spatial data modelling has been the focus of many research projects and different spatial data models are on the market.

The need to share geographic information is well documented ([20], [28], [30]). Research in integration of GIS data sources considering that the barrier of integration is more conceptual than technical in nature. Today, research on interoperability solutions is the way to migrate away from the monolithic systems that dominate the GIS market ([25]).

The first attempts to obtain GIS interoperability involve the direct translation of geographic data from one vendor or standard file format into another. However, these formats translations can lead to information loss. Alternatives that avoid this problem are usually more complex, like standards for spatial data interchange (such as Spatial Data Transfer Standard (SDTS) and Spatial Archive and Interchange Format (SAIF)). A broader discussion of geographic information exchange formats can be found in [12]. One of important strategies for interoperability is conversion of different data formats in common data structure. Such data structure should be forth defined, and usually is based on one of existing GIS standards ([12], [22]).

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 ([20]). But, data standardization is not the whole solution. The interoperability problem would go away if every system always use the same data model to represent the same information (identical names, structure, and representations). OpenGIS standards will do this to some extent, but there are reasons why standardization will not be a complete solution:

– Constructing and maintaining a single, integrated standard data model is difficult problem.

– There will always be a requirement to communicate with information sources that do not conform defined data model standard (legacy systems).

– Existing, legacy information sources have own data models, and there are needs for data conversion from domain model to common model.

– The standard will change, but systems will not all simultaneously change to conform.

Recent reviews of GIS interoperability and integration efforts can be found in [1], [2], [17]. Making local geographic datasets available publicly and establishing a common interoperability framework over shared data interchange protocols are important parts of this research. However, there are institutional and technical problems of geodata sharing and interoperability. These problems have become the focus of international research and infrastructure efforts ([19], [23]). Also, several spatial data interoperability testbeds have been developed – most notably are the Digital Earth Initiative ([8]) and OGC's Web Mapping Testbed ([22]).

A number of proven and well–established methods exist that allow heterogeneous data sources to communicate, including federated databases and schema integration ([16]), object–oriented approaches ([7]), data warehousing ([31]) and mediators and ontologies ([11], [15], [28], [32]). The data warehousing approach ([31]) implies accumulation of spatial data in a few well–defined and tightly connected data stores, where information integration is "pre–computed". While efficient for a relatively small number of core spatial datasets, this approach is not readily extensible to a larger number of datasets with semi structured and ad hoc data. Mediator–based systems, alternatively, 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 ([32]). Structural and syntactic heterogeneity may be solved by mediation.

Mediator–based system is important for spatial data interoperability architecture ([29]). 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 ([32]). A wrapper is a program that is specific to every data source ([4], [24], [27]). Wrapper extracts a set of tuples from source file and performs translation in the data source format. The most important fact is that data integration system lets users focus on specifying what they want rather than thinking about how to obtain the answers. As a result, it frees them of combining data from multiple sources, interacting with each source and finding the relevant sources. Nowadays, mediation concept is a part of the ARPA I3 (Intelligent Information Integration) reference architecture ([4]). 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.

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 use of ontologies as semantic translators is possible approach to overcome the problem of naming conflict and semantic heterogeneity. Recently, the use of ontology in information systems is discussed in [15] and specifically in GIS building in [10], [17], and creation of GIS software components from ontologies in [11]. 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 Interoperability Framework

In this section, we present the *GeoNis* semantic GIS interoperability framework based on mediators and wrappers. 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.

The *GeoNis* framework is based on ORHIDEA ([21], [28]) mediator platform, which uses mediation technology to allow communications between GIS applications over the Internet/Intranet. Applying the mediator framework to the Intranet/Internet environment solves the difficult problem of gaining access to real world data sources. Internet provides the underlying communication layer and protocols for mediation of distributed systems.

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) ([20]) 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.

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 is developed for a particular domain, in our case for GIS applications in local city services, which deals with network data structures (local Telecom with telephone cables, water and soil–pipe services, power supply services, etc.).

3.1. Prerequisite for Interoperability

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. Specified communities have own GIS application, often created with different GIS tools and with different underlying database management systems.

To achieve interoperability, the first prerequisite is that individuals and organizations (i.e. Geographic Information Community - GIC) know each other and the data they possess. Second, there must be a willingness to make data available to users outside the source organization. Given that an organization is open to interoperability, it must announce its existence and willingness to exchange information. Then other individuals can discover the organization and assess whether there may be interest in accessing information. The following six presumptions we defined (modified from [18]):

– *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.

3.2. Architecture of the system

GeoNis interoperability framework uses ORHIDEA 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.

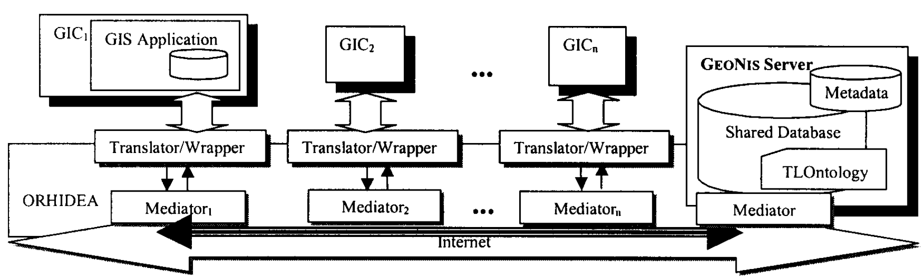

FIG. 1. *GeoNis* interoperability framework architecture

The basic architecture of *GeoNis* framework is shown in Figure 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, in our case that is OGC Simple Feature ([20]). 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) (Figure 2). If it is necessary, query request is forwarded to *GeoNis* server, which resolves problem of semantic heterogeneity in used notation by ontology. *GeoNis* server forwards that demands for data to appropriate data source(s) (local GIS applications). Translator/wrapper on destination GIC application converts request to local application data model (SQL or API). Data source application may execute these requests.

Main demand that should be realized in *GeoNis* platform was to extract data from existing sources of information and to provide data transfer to other applications. However, there is a problem with ontology and data transfer between mediators across the Internet. Extensible Markup Language (XML) ([33]) seems to be ideal solution of this problem. Since data in XML documents are self–defined, user does not need to think about API functions in order to extract data from data sources. This fact is used in *GeoNis* platform in a way that wrappers communicate with concrete information source throw API functions and SQL queries. Wrappers generate ontology as XML document and then forward XML data to mediators. Also, translator converts query results from local application data model to common model. As result, translator/wrapper generates a GML ([14]) document and then forward data to application mediator who asked for them. GIS application, which received data in GML format, could display them or convert them to local GIS format by translator (in keeping with defined styles). The XML document type declaration contains or points to markup declarations that provide a grammar for a class of documents. This grammar is known as a document type definition, or DTD. Mediator uses DTD schema to decode XML document and send it across the Internet. Application gets the needed data in XML format.

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 ac-

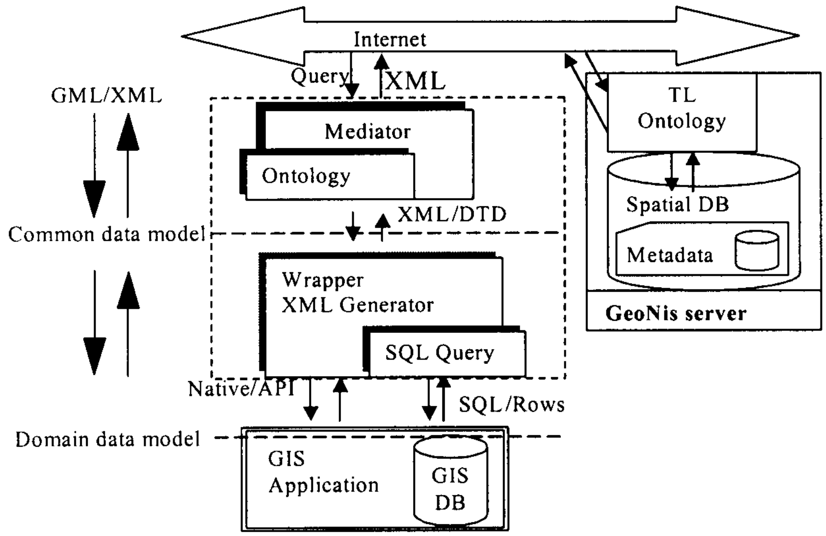

FIG. 2. Data flow in *GeoNis*

cess 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.

3.3. Resolving semantic conflicts in GeoNis

Geodata have geometric and the thematic aspects, both of them with prime importance in terrain description and analysis. To share date between communities required that applications must deal not only with geometric and the thematic aspects, but also with the semantic meaning of spatial data. Semantic characteristics of data depend of their intent in GI community. *Semantic properties* are the definition of the entity or phenomenon. Unlike geometry, which does not vary from one GIC to the next, semantics may vary from group to group in the same way that the conceptual view does. Definitions and meanings may vary slightly or radically, as in the different ways that farmers and civil engineers might define roads ([20]).

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.

The process flow of the data interoperability and semantic conflict resolution in *GeoNis* is:

(1) User (data receiver) sends the query Q through translator,

(2) Local translator send request to the mediator,

(3) 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,

(4) If there is the conflict, the mediator resolve the semantic conflict through semantic process, reformulate and get the right query Q',

(5) Then, the mediator decides where requested data are located, and resend query (queries) to adequate information source(s) (data provider)

(6) Data provider translator accesses the right information from the information provider, converts data to common model and returns the result to the local mediator.

(7) 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).

3.4. ORHIDEA ontologies

In ORHIDEA, the ontology is an explicit specification of some topic. For our purposes, it is a formal and declarative representation, which includes the vocabulary (or names) for referring to the terms in that subject area and the logical statements that describe what the terms are, how they are related to each other, and how they can or cannot be related to each other. Ontology therefore provides a vocabulary for representing and communicating knowledge about some topic and a set of relationships that hold among the terms in that vocabulary.

Our geographic ontology defines geographic objects (municipality borders, administrative regions, blocks, parcels, streets, etc), fields, and spatial (topological) relations. We propose a set of formal definitions that can be used as a basis for algebraic formalization for spatial ontologies. Such ontology can be formalized through definition of classes, relations, functions, and axioms. The ontology together with a set of individual instances of classes constitutes a knowledge base where classes describe concepts in the domain. Although ontology is independent of a particular language, it is necessary to choose a language to describe it. In order to share , exchange, and combine ontologies, the language must be formal. As a implementation language, we used description logic and description logic reasoner.

The ontology class library contains an extensible hierarchy of attributes and an extensible feature hierarchy. The feature hierarchy contains both spatial and non–spatial ontologies. GIC are the local ontology publishers, which develops and maintains domain ontology libraries. Ontologies are defined independently from the actual data, reflect a common understanding of the semantics of the domain of discourse and are used to share and exchange information between sources. They are declarative specifications of the basic notions in a domain. Basic mechanism for expressing semantic interconnectedness is the generalization/specialization classification ("isa" relation).

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

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

2) $R = \{r_i | i = 1, n\}$, is a set of binary typed roles (or relations) between concepts, defined as follows:

$$R = \{(c_1, c_2) | c_1, c_2 \in C\} ,$$

3) *isa* is a set of inheritance relationships defined between concepts, defined as:

$$isa(c_1, c_2) = \{(c_1, c_2) | c_1, c_2 \in C \wedge E(c_1) \subset E(c_2)\} .$$

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*).

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 features. In ORHIDEA ontologies, we define "topological" relationship between concepts which represents topological relations between real world entities. We have defined next "topological" relationship: $T = \{arc-node, \ route, \ node-route, \ point-event\}$, 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.

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.

ORHIDEA uses *isa* relation to define a specialization relation (or hyponyms) to establish a hierarchical *taxonomy* of concepts. Relations in ontology definitions must be defined independent by concepts. In the following, we show how Description Logic ([9]) can represent ontologies by such features.

An example of a primitive concept (from the Telecom domain) definition is as follows:

```
(defconcept Cable_Segment (?r Telephone_Cable)
   :=> (and (exists (?j)
            (and (Junction ?j)
```

```
            (= (starts_at ?r) ?j)))
        (and (exists (?k)
            (and (Junction ?k)
                (= (ends_at ?r) ?k)))
        (= (dimension ?r) LINEAR)))
```

It defines a primitive concept "*Cable_Segment*" which is a sub-concept of "*Telephone_Cable*". It also states that all "*Cable_Segment*" have at least one "Junction" with which they have "starts–at" relation, and at least one "Junction" with which they have "ends–at" relation.

The non–primitive concept "*Local_Road*" (concept from Road Network database) is defined as follows:

```
(defconcept Local_Road (?r Road)
    :<<=>> (< (width ?r) 20))
```

The concept "*Local_Road*" is a subconcept of "*Road*" and its "*width*" is filled by a value less than 20. Therefore, every instance of "*Road*" whose "*width*" relation is filled by a value less than 20 will be classified as "*Local_Road*" and vice versa.

Between concepts in taxonomy may be defined primitive relations. Here is an example of a primitive relation definition:

```
(defrelation dimension
    ((?se Spatial_Element) (?sd Spatial_Dimension)))
(defrelation bounds ((?x Spatial_Element)(?y Spatial_Element))
    :=> (and (dimension ?x POINT)
            (dimension ?y LINEAR))
(defrelation starts_at
    ((?x Spatial_Element)(?y Spatial_Element))
    :=> (bounds ?y ?x))
```

Relation "*starts_at*" is defined as *bound* (relation *bounds*) between two spatial elements. One of them is with *POINT dimension*, and another is with *LINEAR dimension* (relation between point and line spatial elements).

3.5. Semantic inter–correspondences between ontologies

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.

In ORHIDEA we divide the semantic conflict (semantic inter–correspondences) into four types:

– 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 follows:

$$SEqu(c_1, c_2) = \{(c_1, c_2) | c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) = E(c_2)\} .$$

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 follows:

$$\begin{aligned} &SNEqu(c_1, c_2) \\ &= \{(c_1, c_2) | c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) \neq E(c_2) \wedge Name(c_1) = Name(c_2)\} . \end{aligned}$$

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 follows:

$$\begin{aligned} &SIntersec(c_1, c_2) \\ &= \{(c_1, c_2) | c_1 \in O_1 \wedge c_2 \in O_2 \\ &\wedge E(c_1) \cap E(c_2) \wedge E(c_1) \not\subset E(c_2) \wedge E(c_2) \not\subset E(c_1)\} . \end{aligned}$$

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 follows:

$$\begin{aligned} &SContain(c_1, c_2) \\ &= \{(c_1, c_2) | c_1 \in O_1 \wedge c_2 \in O_2 \wedge E(c_1) \not\subset E(c_2) \wedge E(c_2) \subset E(c_1)\} . \end{aligned}$$

This kind of relation is neither commutative nor transitive.

Relationships between reference (common) model object classes and application ontology classes (concepts) define the semantic of a geographic data set. The basic semantic relationship, abbreviated as *Refers_to*, is between concepts from local ontology and real–world classes (represented in common model). It is defined by predicate: *Refers_to(RefClass a, OntologyConcept c)*. For example, such relationship exist between concept *Cable* (from Telecom domain) and reference model object class *Polyline: Refers_to(Polyline, Cable)*.

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:

$$\begin{aligned} & SRelev(b_1, c_1) \\ = & \left\{ (b_1, c_1) \in B \times C | \exists a \in A \wedge Refers\_to(a, b_1) \wedge Refers\_to(a, c_1) \right\} . \end{aligned}$$

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.

4. Conclusion

The paper presents an ongoing case study and development of framework for semantic interoperability between heterogeneous GIS data sources. This framework is aimed to resolve interoperability problem in local, municipality environment. The project introduces a mediation–based system designed to allow data source reusability, system scalability and extendibility. 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. System should provide actualization of client/server applications, using Internet and Web technologies as under-layer for network service and integration of distributed data sources.

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. Mediators should hide existence of various (heterogeneous) data sources. The mediator and agent-wrapper approach also allows maintaining these applications (data sources) and incorporate new sources, as they became available ([26]). The mediators provide coherent views of the data in the repositories by performing semantic reconciliation of the data representations provided by the wrappers.

Mediators are efficient and cheap way of integrating data from heterogeneous information systems. Logical consequence is that mediators provide evolving system development and use existing software investments in software and databases. Mediators are not just simple interface between applications and databases but they have to include some knowledge in themselves that cannot exist in data they work on. They have to aggregate underlying data depending on criteria dictated by the application layer. The process of understanding which domain contains the best information for answering a query is delegated to the automatic knowledge based engines built in the mediators themselves. Mediators are a good way for integrating completely different data source types.

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 modelling 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.

GeoNis would also work with other types of data and another application domains, but with certain changes. Replacing of OGC common data model could perform such generalizations of the system. Also, there is need for changes in implementation of translators (or realization of new translators). The large problem for the design of GeoNis, irrespective to application domain, is ontology development. Such process depends on knowledge and contribution of domain experts. Recognition/definition of semantic intercorrespondences and fuzzy mapping between domain ontologies is additional problem. That is long and hard work that demands contribution of experts from different domains and commitment about concepts meaning. Our approach predicts interactive user (or domain expert) participation in definition of new rules. If interpretation of received data was not correct (i.e. wrong mapping between ontologies), user generates rule that is used latter for generating new semantic intercorrespondences.

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 work was supported in part by the Serbian Ministry of Science, Technology and Development under Contract No. IT.1.23.0249A (project: *"Geographic Information System for Local Authorities based on Internet/WWW Technologies"*).

REFERENCES

1. D. J. Abel et al.: *Towards Integrated Geographical Information Processing.* International Journal of Geographic Information Science **4** (1998), 353–371.
2. J. Van den Akker and A. Siebes: *A Data model for autonomous objects.* Computer Science/Department of Algorithmics and Architecture, CWI Technical Report, CS-R9539, Amsterdam, Netherlands, 1995.
3. J. Van den Akker and A. Siebes: *DEGAS: A temporal active data model based on object autonomy.* Computer Science/Department of Algorithmics and Architecture, CWI Technical report, CS-R9608, Amsterdam, Netherlands, 1996.
4. 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).
5. Y. A. Bishr, H. Pundt, W. Kuhn and M. Rdwan: *Probing the Concepts of Information Communities – A First Step Towards Semantic Interoperability.* In: Interoperating Geographic Information Systems (M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, eds.), Kluwer Academic Publishers, Norwell, MA, 1999, pp. 55–70.
6. Y. A. Bishr: *Overcoming the Semantic and Other Barriers to GIS Interoperability.* International Journal of Geographic Information Science **4** (1998), 299–314.

7. S. Chawathe, H. Garcia–Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ulman and J. Widom: *The TSIMMIS Project: Integration of Heterogeneous Information Systems.* In: The Proceedings of ISPJ Conference, Tokyo, Japan (October 1994), pp. 7–18.
8. Digital Earth Initiative: 2002. http://www.digitalearth.gov.
9. KRSS Group of the ARPA Knowledge Sharing Effort: *Description–Logic Knowledge Representation System Specification.* http://dl.kr.org/krss-spec.ps (2002).
10. T. Devogele, C. Parent and S. Spaccapietra: *On Spatial Database Integration.* International Journal of Geographic Information Science **4** (1998), 335–352.
11. F. Fonseca and M. Egenhofer: *Ontology–driven information systems.* In: 7th ACM Symposium on Advances in GIS (C.B. Medeiros, ed.), Kansas Sity, MO, 1999, pp. 14–19.
12. Diffuse Project: *Geographic Data Exchange Standards.* http://www.diffuse.org/gis.html (2001).
13. M. Genesereth et al.: *Infomaster: An Information Integration System.* ASIGMOD Conference, 1997, pp. 539–542.
14. Open GIS Consortium: *GML 2.0 Recommendation paper.* http://www.opengis.net/gml/01-029/GML2.html (2001).
15. N. Guarino: *Formal Ontology in Information Systems.* In: Formal Ontology in Information Systems. Proceedings of FOIS'98 (N. Guarino, ed.), Trento, Italy (6–8 June 1998), IOS Press, Amsterdam, pp. 3–15.
16. A. Sheth and L. Larson: *Federated database systems for managing distributed, heterogeneous and autonomous databases.* ACM Computing Surveys **22** (1998), 183–236.
17. R. Laurini et al.: *Spatial multi–database topological continuity and indexing: a step towards seamless GIS data interoperability.* International Journal of Geographic Information Science **4** (1998), 373–402.
18. A. Levinsohn: *Geospatial Interoperability: The Holy Grail of GIS.* GeoEurope, available online, http://www.geoplace.com/gw/2000/1000/1000data.asp.
19. I. Masser: *Governments and Geographic Information.* Taylor & Francis, London, 1998.
20. L. McKee and R. Buehler: *The Open GIS Guide.* Third Edition, OpenGIS Consortium, Inc. (1998), http://www.OpenGIS.org/techno/specs.htm.
21. D. Mitrović, A. Mitrović and S. Đorđevic–Kajan: *A scalable, object–oriented GIS Framework.* In Proc. of ISPRS Int. Workshop "Hardware and Software Aspects of GIS", Milan, Italy, 1996, pp. 20–21.

22. Open GIS Consortium: *Web Mapping Testbed Public Page.* http://www.opengis.org/wmt (2000).
23. H. Onsrud, G. Rushton: *Sharing Geographic Information.* Rutgers University Press: New Brunswick, NJ, 1995.
24. M. T. Roth and P. Schwarz: *Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources.* In: Proceedings of the 23rd VLDB Conference, Athens, Greece, 1997, pp. 226–275.
25. M. Sondheim, K. Gardels and K. Buehler: *GIS Interoperability.* In: Geographical Information Systems 1 Principles and Technical Issues (P. Longley, M. Goodchild, D. Maguire, and D. Rhind, eds.), John Wiley & Sons, New York, 1999.
26. L. Stoimenov, A. Mitrović and S. Đorđević–Kajan: *An object–relational environment for developing object–oriented GIS applications based on mediator.* In: 16th IASTED International Conference on Applied Informatics, Garmisch–Partenkirchen, Germany (February 1998), pp. 168–171.
27. L. Stoimenov, A. Mitrović, S. Đorđević–Kajan and D. Mitrović: *Bridging objects and relations: a mediator for an OO front-end to RDBMSs.* Information and Software Technology, Elsevier **2** (1999), 59–68.
28. L. Stoimenov, S. Đorđević–Kajan and D. Stojanović: *Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology.* In: MELECON 2000, Cyprys, 2000.
29. L. Stoimenov, S. Đorđević–Kajan and A. Milosavljević: *Realization of Infrastructure for GIS Interoperability.* In: YUINFO '02, Kopaonik, 2002 (Serbian).
30. A. Vckovsky (ed.): *International Journal of Geographic Information Science – Special Issue: Interoperability in GIS.* **4** (1998).
31. A. Voisard and M. Juergens: *Geographic Information Extraction: Querying or Quarrying?.* In: Interoperating Geographic Information Systems (M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, eds.), Kluwer Academic Publishers, New York, 1998.
32. G. Wiederhold: *Interoperation, Mediation and Ontologies.* Int. Symp. on Fifth Generation Computer Systems, Tokyo, Japan, 1994.
33. World Wide Web Consortium: *XML, Extensible Markup Language.* http://www.w3.org/TR/REC-xml (2002).

University of Niš
Faculty of Electronic Engineering
Department of Computer Science
P. O. Box 73, 18000 Niš
Yugoslavia