Intelligent Discovery and Retrieval of Geoinformation using Semantic Mediation and Ontologies

Aleksandar Stanimirović[1], Leonid Stoimenov[2], Slobodanka Đorđević-Kajan[3]

Abstract-**In this paper we proposed architecture, based on modified hybrid ontology approach and semantic mediation, for solving the problems of semantic heterogeneity. Process of semantic mediation is based on ontologies that provide abstract description of the content of information sources. It also uses ontologies to provide semantic consistency and geodata exchange with terminology translation. Proposed architecture provides means for building intelligent systems that can integrate a number of different information sources.**

Keywords-**semantic heterogeneity, ontologies, semantic mediation, geoinformation retrieval, geoinformation discovery**

I. INTRODUCTION

In many areas of life, geographic information is key factor in planning and decision-making. In such environment, accurate and simple methods for identification and retrieval of, often distributed and heterogeneous, geodata is very crucial. Interoperability of GIS (Geographic Information System) is a possible solution for these problems.

Interoperability of information systems relies on bases of agreement that describe what is shared among information sources. Interoperability means openness in the software industry, because open publication of internal data structures allows GIS users to build applications that integrate software components from different developers. Interoperability also means the ability to exchange data freely between systems, because each system would have knowledge of other systems formats [1].

Nowadays, providing a user with huge amount of data from different sources across networks or inter-networks in a short amount of time is done by interoperable systems. Users of spatial data expect information systems to help them to search and process the data as well as supporting them with necessary information and knowledge about the data – i.e., metadata. They also expect to have homogenous interfaces for managing, modeling and processing the data from different sources. Though interoperability has to overcome complexity of conversion and integration process, there is a long way from data transfer and data format exchange (or conversion) to system interoperability. Interoperability issues not only refer to different structures and models of data sets, but also to the different methods and operations applying to the data.

In the domain of GIS interoperability, differences in data sources, disciplines, tools and repositories can cause heterogeneity. Interoperability helps to reuse geodata and avoid waste of our assets. It has the potential to offer a prompt reaction to obtain suitable data set hen dealing with geographic information analysis (such as natural disaster management). In order to achieve the sharing and exchange of existing data between different departments, semantic heterogeneity between data from different component systems must be taken into account. Without knowing the semantics of data (i.e. their meaning), it is impossible to integrate and understand data appropriately. Unfortunately, there are only few integration approaches that consider the semantics of data.

In this paper we focus on our approach for discovering and interpretation of geographic information based on ontologies. The rest of paper is structured as follows. In section II we discuss heterogeneity problems during discovery and retrieval of geographic information. Section III describes GeoNis generic architecture for interoperability and the role of semantic mediators in it. Section IV describes our architecture for semantic mediation based on ontology approach.

II. RELATED WORK

Geographic data set integration is the process of establishing relationships between corresponding object instances in different, autonomously produced, geographic data sets of a certain domain. The purpose of geographic data set integration is to share information between different geographic information sources. Geographic data set integration gets more and more attention nowadays since the digitizing of traditional map series has ended. In these map series, corresponding object instances were only linked implicitly by a common spatial reference system. In order to make these relationships explicit geo-science researchers and computer scientists have developed various strategies.

One important initiative to achieve GIS interoperability is the OpenGIS Consortium [2]. 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. But, data standardization cannot solve the whole problem. 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). OpenGIS standards will only partially solve this problem.

Mediator-based system is important for spatial data interoperability architecture [3]. The 3-level architecture of mediator-based systems is constructed from an application layer, and large number of relatively autonomous information sources (heterogeneous data sources with wrappers), communicating with each other over a standard protocol [4]. A wrapper is a program that is specific to every data source [5]. Wrapper extracts a set of records from source file and

[1]Aleksandar Stanimirović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: alex@elfak.ni.ac.yu
[2]Leonid Stoimenov is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: leni@elfak.ni.ac.yu
[3]Slobodanka Đorđević-Kajan is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: sdjordjevic@elfak.ni.ac.yu

performs translation in the data source format. The most important advantage is the fact that data integration system allows users to 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 [6].

The uses of ontologies as semantic translators is approach that can possible overcome the problem of naming conflict and semantic heterogeneity. 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, and information systems [7].

During past several years many different solutions for problems of semantic heterogeneity have been proposed [8, 9, 10]. In this paper we proposed our approach for discovering and retrieving geodata based on ontologies and semantic mediation.

III. GeoNis Framework for semantic interoperability

GeoNis is a generalized framework for GIS interoperability. It provides infrastructure for data interchange in the local community environment [3, 11]. The basic architecture of GeoNis framework is shown on Fig. 1.

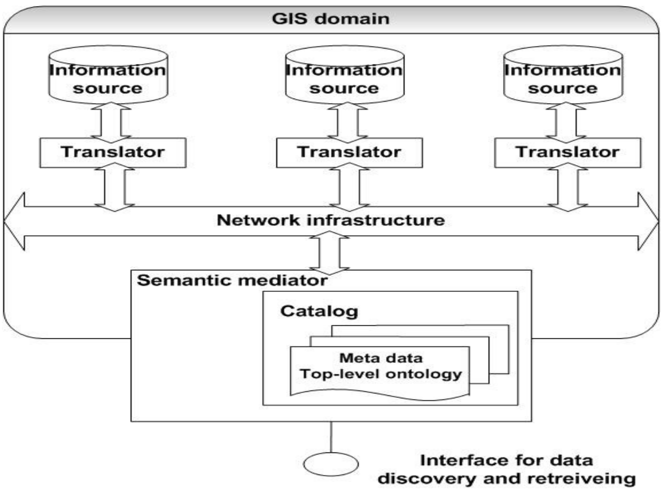

Fig. 1. GeoNis – framework for semantic interoperability [12].

Generic architecture of GeoNis recognizes several different components that have important role in geoinformation discovering and retrieving process [13]:

- *Information source* – in each node of GeoNis framework there exist GIS application and corresponding (spatial and non-spatial) data sources. Data in these local data sources are accessible according to user privileges.
- *Translator* - component that translates information flow between information source and GeoNis system.
- *Semantic mediator* - requests for specific data set are forward through.
- *Catalog* – maintains metadata and all shared/common geographic data as addition to domain oriented GIS applications.

When user (human or GIS application) issues a request, system have to decide which source (or sources) is able to deliver the requested information. The simplest way to do that is to query each information source that is registered. If the request succeeds, the results are delivered to the requester. In the sense of intelligent information discovery and retrieval such behavior is not aspired. We want the system to be aware which sources are worth of querying and which are not. This goal can be reached by providing an abstract description of the content of an information sources using ontologies.

Architecture that we propose in this paper (Fig. 2) solves the problems of geoinformation discovery in distributed and heterogeneous environment. Proposed architecture is based on generic Semantic Mediator architecture given in [10]. Proposed architecture must provide techniques for realization of domain oriented mediator and translator chaining in a specific domain. This architecture also provides means that support interpretation of retrieved geoinformation. Ontologies have a central role in resolving semantic conflicts.

All access to geoinformation in local community environment goes through *Web Feature Service* defined by proposed OGC technology standards [2]. We enhanced this interface with additional functionality in order to support user profiles and privileges. This interface is implemented by *Semantic Mediator* component. This component acts as an access point for a number of independent geoinformation sources and allows integration of their information bridging over the semantic differences among them. *Semantic Mediator* mediator enables users to access multiple information systems as though they were a single system with a uniform way to retrieve information and perform computations. It accepts high-level requests from users and automatically translates them into a series of lower-level requests for different GICs. Results of this low-level request are then combined into a result for a high-level request. In order to accomplish this task Semantic Manager must use metadata information, provided by *Ontotology Manager* services, in order to discover every concept (top-level or local) in environment that can be targeted with user request.

Ontology Manager is a component that provides access to the shared metadata that resides on the common server. This component implements interfaces that provide means for discovery, access and management of metadata to the rest of the community. Interface *Publish service* allows GIC nodes to register their data and services and to create relations between their local ontologies and top-level ontologies. Interface *Query service provides* functionality for locating and accessing requested metadata.

Intelligent UI component dynamically generates web-based user interface according to user privileges and user profiles. This interface allows users to query local community for metadata and geodata according to their privileges.

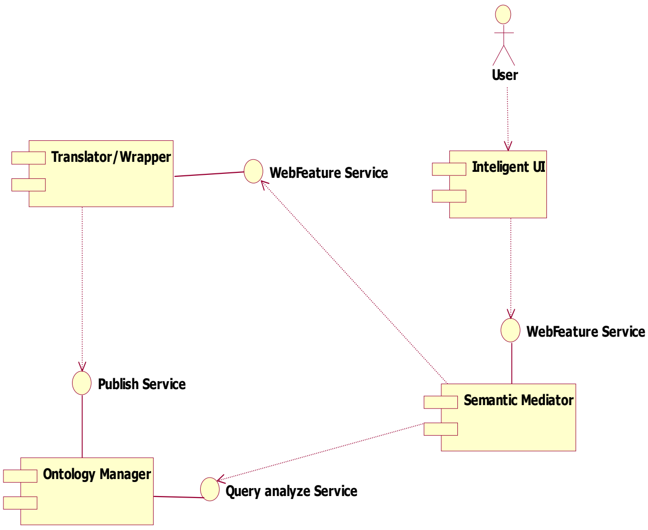

Fig. 2. Components and interfaces for semantic mediation

IV. Semantic Mediation

In GeoNis Semantic Mediator we propose a semantic based integration approach that uses multiple ontologies, instead of an integrated view. In this context, ontologies are virtually linked by inter-ontology relationships, which are then used to indirectly support query processing. Our Semantic mediator uses hybrid ontology approach (Fig. 3). Meaning of the terminology of each community is specified in the local ontology. Semantic Mediator provides a methodology and software support for semantic mismatches (conflicts) resolving between terminologies. This methodology uses the ontology mappings between each community terminologies and a top-level ontology or the common data model (reference ontology).

All meta-data information (top-level ontologies, inter-ontology relations, ontology mapping rules and additional metadata information) in local community environment are hosted by Shared server. This metadata describes the properties of geo-data sources that can be queried through mediators and translators. Using meta-data from Shared server, as respond to a user query, Semantic Mediator can return sets of geo-objects (features) from different data sources. Contained public information can be classified as follows:

- *User metadata*:
    - *User privileges* – user rights for accessing data in local GIC nodes or in common GeoNis server.
    - *User profiles* – description of customized, intelligent, web-based user interface that is dynamically generated whenever user access data in local community environment.
- *Virtual organization metadata (Ginis Catalog)* – local community structure description. Every new GIC who wants to participate in exchanging data must register with common GIS server in order to allow access to his public available data and local ontology. After that, users from registered GIC have access to all available data from other public GIC databases and access to shared data owned by shared GIS server (with possible given rights for access).
- *Top-level ontologies* – domain shared vocabulary and description of available data sets
- *Ontology mappings* – describes mapping between top-level and local (application) ontologies.

When user wants to retrieve some information from local community environment, first, he has to logon on the system. According to his user profile and privileges *Intelligent UI* component builds appropriate Web based interface. This interface also includes tools that can help user to build query using top-level concepts obtained from local community environment.

After logon procedure, user can query environment for geoinformation that he is interested in. In order to provide requested data, *Semantic Mediator* must analyze user query and discover every concept (feature) in environment that fits to user request.

This analysis is done by *Ontology Manager*. User query is first matched against top-level ontologies (description of the available datasets in information sources). In a highly heterogeneous system it might happen that the representations of queries are incompatible with some top-level ontologies, because they use a different terminology or different structuring principles for the domain. In such case rules for conflict resolving must be used (ontology mappings).

Ontology Manager also has to deal with distributed geospatial datasets. That means that user request can hit several information sources at the same time and a result data set is not located in one data source. Columns that forms result data set can be distributed in great number of data sources. Also, result data is not necessary in relational databases. Spatial data can be stored in great number of heterogeneous data sources. *Ontology Manager* builds *FeatureSchema*, metadata information that contain description and location of every concept (feature) in environment that can be treated as a result of a user query [14]. Result of user query is treated as a collection of feature tables and spatial reference systems definitions. Every feature table description contains basic table meta-data information and a collection of different connections. Feature table connections describe information sources that have to be used to build up feature table data. These connections also contains information of means for retrieving and rules for integrating data from different sources. Using this schema information, *Semantic Mediator* retrieves data from local GIC nodes and builds result dataset that is forwarded to the user, as a response to his query, in the form of a GML document.

V. Conclusion

In this paper we introduced generic architecture and framework for ontology based discovery and retrieval of geoinformation. This architecture provides us with means for building intelligent system that can integrate a number of different information sources bridging over syntactic and semantic differences between them. Intelligence of this system is based on techniques for semantic mediation.

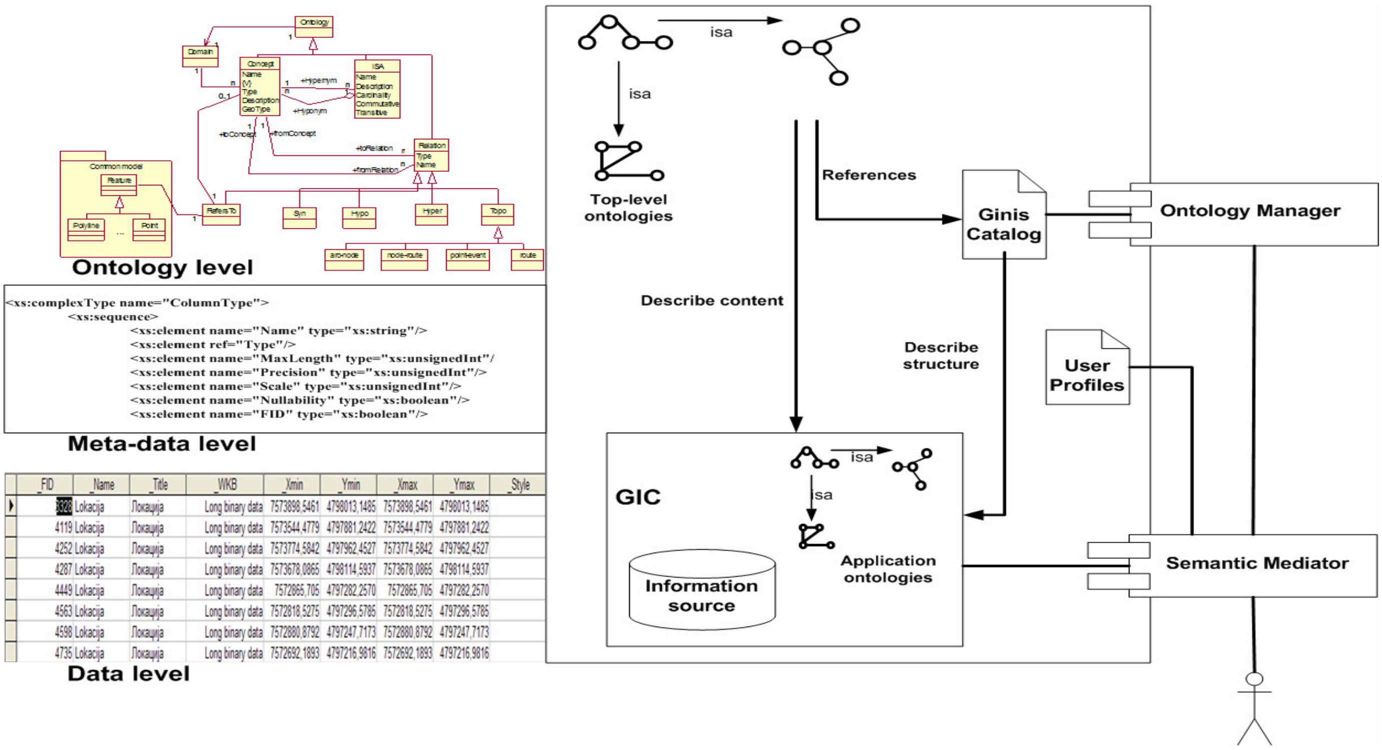

Fig. 3. Hybrid ontology approach

Semantic Mediator acts as an access point for a federation of independent geoinformation sources. When processing user requests, Semantic Mediator must be aware, in advance, which information sources contain data that fulfill user requests? This goal can be reached by providing an abstract description of the content of an information sources using ontologies. It also uses domain ontologies in order to provide semantic consistency for data from that domain and geodata exchange with terminology translation.

Proposed architecture is a component-based with strongly defined interfaces between components. In this way architecture can be easily extended in various directions. Introduced architecture solves the problems of semantic heterogeneity in a single GIC environment domain.

References

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

[2] Open Geospatial Consortium, http://www.opengeospatial.org , 2005

[3] Stoimenov L., Djordjević-Kajan S., Stojanović D., 2000, "Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology", *MELECON 2000, 10th Mediterranean Electrotechnical Conference*, May 29-31 2000, Cyprus, Proceedings Vol. 1, pp 334-336.

[4] Wiederhold G., "Interoperation, Mediation and Ontologies", *Int. Symp., on Fifth Generation Computer Systems*, Tokyo, Japan, 1994.

[5] Stoimenov L., Mitrovic A.,Djordjevic-Kajan S., Mitrovic D., "Bridging objects and relations: a mediator for an OO front-end to RDBMSs", *Information and Software Technology*, Elsevier, 1999, Vol 41, No. 2, pp. 59-68.

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

[7] Guarino N., "Formal Ontology in Information Systems", *In N. Guarino (ed.) Formal Ontology in Information Systems. Proceedings of FOIS'98*, Trento, Italy, IOS Press, Amsterdam, 1998, pp.3-15.

[8] Bernard L., Einspanier U., Haubrock S., Hübner S., Kuhn W., Lessing R., LutzM., Visser U., 2003, "Ontologies for Intelligent Search and Semantic Translation in Spatial Data Infrastructures", *Photogram-metrie - Fernerkundung - Geoinformation*, 2003, pp. 451-462.

[9] Klien E., Einspanier U., Lutz M., Hübner S., 2004, "An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information", *Proceedings of the 7th Agile Conference on Geographic Information Science*, Heraklion, Crete, 29 April-1 May, 2004

[10] Stoimenov L., Stanimirović A., Đorđević-Kajan S., "Semantic Interoperability using multiple ontologies", *Proceedings of 8th AGILE Conference on GIScience,* Estoril, Portugal, May 26-28, 2005, pp.261-269.

[11] Stoimenov L., Djordjević-Kajan S., 2002, "Framework for semantic GIS interoperability", *Journal FACTA Universitatis, Series Mathematics and Informatics*, 17(2002), pp.107-124.

[12] Stoimenov L., Đorđević-Kajan S., 2003, "Realization of GIS Semantic Interoperability in Local Community Environment", *Proceedings of 6th AGILE conference on Geographic Information Science: "The Science behind the Infrastructure"*, ISBN 2-88074-541-1, AGILE 2003, Lion, France, April 20-23.2003. pp.73-80

[13] Stoimenov L., Stanimirović A., Đorđević-Kajan S., "Development of GIS Interoperability Infrastructure in Local Community Environment", *From Pharaohs to Geoinformatics, FIG Working Week 2005 and GSDI-8 Cairo*, Egypt April 16-21, 2005, TS41.2.

[14] Stanimirovic A., Stoimenov L., Đorđević-Kajan S., "Geodata integration in distributed environment using OLE DB data provider", *ICEST 2004*, Bitola, Jun 2004, pp. 329-332