Semantic Interoperability Using Multiple Ontologies

Leonid Stoimenov, Aleksandar Stanimirovic, Slobodanka Djordjevic-Kajan
CG&GIS Lab, Department of Computer Science
Faculty of Electronic Engineering, University of Niš, Serbia and Montenegro
E-mail: leni@elfak.ni.ac.yu, alex@elfak.ni.ac.yu, sdjordjevic@elfak.ni.ac.yu

SUMMARY

There are now robust technical solutions to interoperate geographic information in a syntactic and schematic sense. But, problem with different meaning of geo-data from distributed geo-information sources still exists. The use of ontologies in semantic translation is a viable approach to overcome the problem of semantic heterogeneity. Also, from the technological point of view, mediator-based systems are important for the spatial data interoperability architecture. Our solution for semantic heterogeneity in an interoperable environment (called GeoNis) is a Semantic Mediator. Semantic heterogeneity is resolved by using a hybrid ontology approach. This paper presents architecture of Semantic Mediator. Also, paper defines concepts mappings between community terminologies (local ontologies) and mappings between local ontology and top-level ontology

KEYWORDS: *ontology, interoperability, data semantics, semantic mediator*

INTRODUCTION

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 (Stoimenov, 2000).

Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged (Abel, 1998, Laurini, 1998). In short, information sharing (and information exchange) not only needs to provide full accessibility to the data, it also requires that the accessed data may be processed and interpreted by the remote system. Problems that might arise due to heterogeneity of the data are already well known within the distributed database systems community (Genesereth, 1997).

Great number of independent geodata producers, accessible by means of World Wide Web, has only increase problems of heterogeneity. To enable interoperability, remote systems must be able not only to locate and access data sources, but also to interpret and process retrieved data. In order to achieve this, remote systems had to deal not only with syntactically heterogeneous data objects (objects that are organized following different conceptual schemas) but as well with semantically heterogeneous objects (objects that that have different meaning) (Bishr, 1999, Kuhn, 2002, Stoimenov, 2004a). This has become more important due to the fact that spatial data modeling has been the focus of many research projects and different spatial data models are on the market. Semantics are a key factor of successful interoperability between GI systems.

In this paper we focus on our approach for discovering and interpretation of geographic information based on ontologies. The paper is structured as follows. Section 2 describes related work on interoperability, especially semantic interoperability of GIS and role of mediators and ontologies. In Section 3 we discuss an architecture of semantic mediators and their role as solution of semantic interoperability problem. We also discuss types of ontology mappings between concepts from local ontologies and concepts from global terminology (top-level ontology).

RELATED WORK

Today, research on interoperability solutions promises a way to move away from the monolithic systems that dominate the GIS market. Recent reviews of GIS interoperability and integration efforts can be found in (Laurini, 1998). Also, one important initiative to achieve GIS interoperability is the OpenGIS Consortium (OGC) (OGC, 2004).

The realization of interoperable GISs is difficult process, due to two main systems characteristic - distributed data sources and their heterogeneity (Genesereth, 1997). There are now robust technical solutions to interoperate geographic information in a syntactic and schematic sense (e.g. OpenGIS Consortium), but still exists problem with confusion in the meaning associated to geo-data (semantic heterogeneity). Stuckenschmidt (2000) gives an introduction about problems concerning the syntactic, structural and semantic integration.

From the technological point of view, mediator-based systems are important for the spatial data interoperability architecture (Stoimenov, 2002). Such 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 (Wiederhold, 1998). Structural and syntactic heterogeneity may be solved by mediation.

The importance of semantics in geographic information is well documented (Kuhn, 2002, Egenhofer, 2002). Semantics refers to user's interpretation of the computer representation of the world – i.e., the way users relate computer representation to the real world (Meersman, 1995). In order to achieve semantic interoperability in a heterogeneous information system, the meaning of the information that is interchanged has to be understood across the systems. Systems must be able to exchange data in such a way that the precise meaning of the data (i.e. semantic) is readily accessible and the data itself can be translated by any system into a form that it understands.

The use of ontologies as semantic translators is a viable approach to overcome the problem of semantic heterogeneity (Hakimpour, 2001, Fonseca, 2002, Stoimenov, 2003). An ontology consists of logical axioms that convey the meaning of terms for a particular community. Logical axioms are defining concepts and their relations, and also express constraints on both concepts and relations. An ontology exists under a consensus by members of a community (Bishr, 1999), e.g., users of single information system or people in one discipline.

OGC also paid attention to semantics issues of GIS. There is an interest group under OGC and a draft standard on semantics. Since the Semantics and Community Metadata are the bases of an information community, to transfer a data set from an information community to another OGC suggests the use of something called Semantic Translator.

THE ROLE OF SEMANTIC MEDIATORS IN GEONIS

GeoNis is framework for interoperability of GIS applications that have to provide infrastructure for data interchange in the local community environment (Stoimenov, 2002). Data sources are local services and offices that own geodata in some format. Semantic interoperability in GeoNis, resolved by ORHIDEA components, is the ability of sharing geospatial information at the application level, without knowing or, understanding terminology of other systems.

GeoNis generic architecture (Stoimenov, 2005) recognizes several different components that have important role in geoinformation discovering and retrieving process:

- GIC – every GIC contain GIS application and corresponding spatial data sources.
- Wrapper/translator - each type of information source requires a wrapper/translator that translates information flow between information source and GeoNis system.
- Data in local databases - accessible according to user privileges.

- Semantic mediators - requests for specific data set are forward through.
- Shared GIS server (Catalog Server) - maintains metadata and all shared/common geographic data as addition to domain oriented GIS applications.

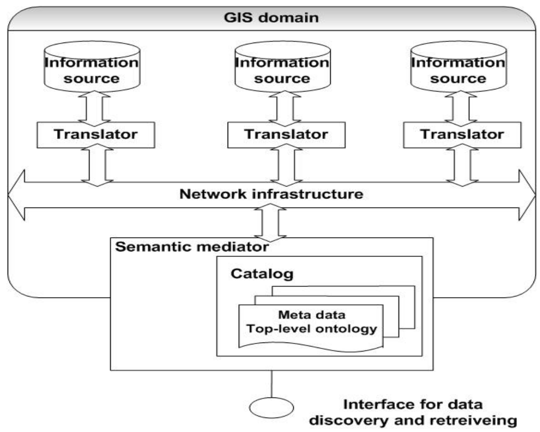

Figure 1. The role of domain oriented mediator in GeoNis architecture

According to GeoNis architecture every GIC environment can have several translators (figure 1). There can be one translator for every information source in GIC environment. Translators can be implemented using different approaches and technologies. Ginis OLE DB data provider is an example of translator implementation (Stoimenov, 2004b). This approach (unified methods for data access) allows simply chaining of translators. Also we can easily add new information sources without influence on other GIC environments. In order to do this we have to provide semantic mapping only for the GIC environment with new information source. In this way we have simplified the problem of semantic heterogeneity.

The basic components providing interoperability (including the resolution of semantic conflict and semantic interoperability) in GeoNis framework are Semantic Mediators. Semantic Mediator architecture and its role in GeoNis framework is given on figure 2 (more detail about GeoNis Mediator architecture you can found in (Stoimenov, 2005)). Basic components of Semantic Mediator for semantic conflict resolution are:

- SemanticMapper components, which performs mapping between semantically similar data, and
- SchemaMapper components, which performs query distribution and reformulation.

The main part of Semantic Mediator is MediatorManager. The basic function of this component is to manage the work of other components. It receives the data produced by other mediators (using DataTranslator), initiates the search for the appropriate translator, receives the processed data from the translator and sends the data to the corresponding mediator (using DataTranslator). TranslatorSelector is the component of Mediator that provides communication with the translators. TranslatorSelector sends (through DataSender) / receives (through DataReceiver) data to/from the MediatorManager.

MediatorManager initiates the system components which resolve semantic conflicts. Semantic Mappings components RuleEngine, OntoManager and SDManager (Spatial Data Manager) participate in resolving semantic conflicts. These components can be parts of the mediator or a distinct component that will communicate with the mediator. The components use the meta-data, schemas and knowledge from GinisCatalog.

The GinisCatalog is a collection of metadata concerning to data stored in a distributed geo-data sources (geographic databases). This metadata describes the properties of geo-data sources which can be queried through mediators and translators (namely ORHIDEA Mediator platform). The GinisCatalog interface provides operations to retrieve semantic mappings between concepts from local ontologies, the geo-object types that it can handle, and schema definition of a feature type. Using meta-data from GinisCatalog, as respond to a user query, Semantic Mediator can return sets of geo-objects (features) from different data sources. Figure 3 shows meta-model for Ginis Catalog ontology repository as an UML class diagram.

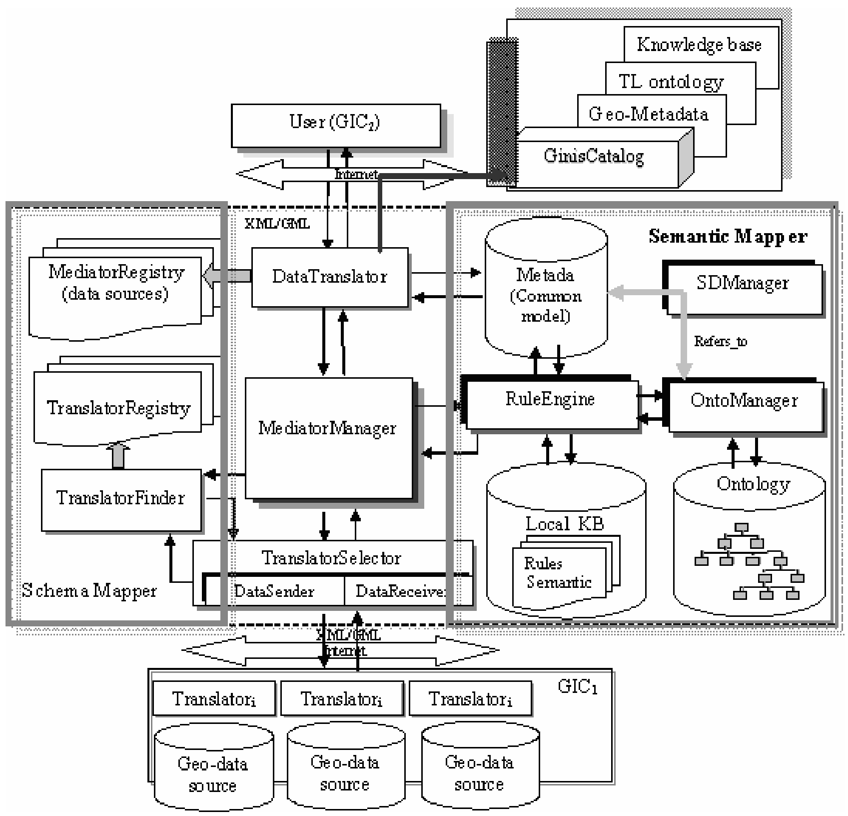

Figure 2: Architecture and role of Semantic Mediator in GeoNis.

GeoNis architecture doesn't clearly define the role of semantic mediators and relations between semantic mediators and GICs. Let presume that we have a semantic mediator for a specific domain (for example, applications in local city services, which deal with network data structures such as Telecom, water and soil pipe services, power supply services, and some local government services). In such case we can threat semantic mediator as an access point for discovering and retrieving semantically heterogeneous data in that domain (*Figure 4*). If there are metadata and top-level ontology (share vocabulary) for a domain provided to a semantic mediator, it can provide semantic consistency for data from that domain and geodata exchange with terminology translation as described in GeoNis architecture.

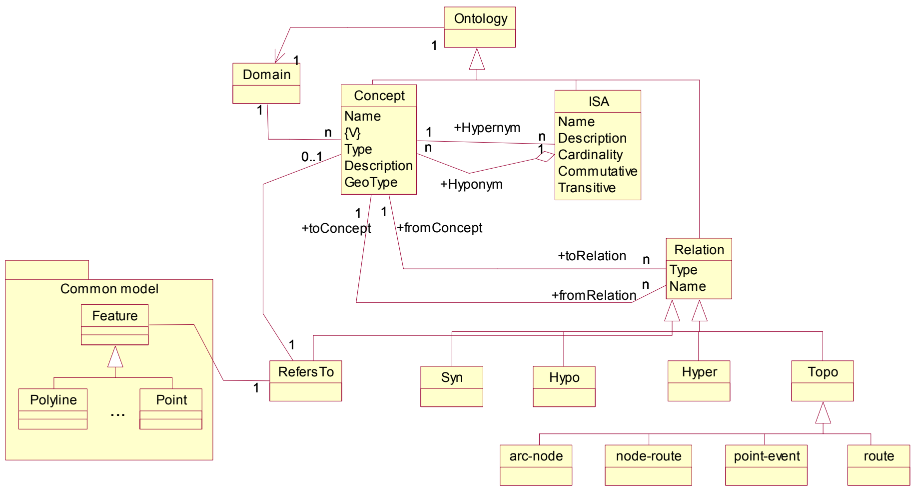

Figure 3: Meta-data ontology repository.

Generic GeoNis architecture also allows chaining of the semantic mediators. Metadata and top-level ontology must be provided in order to integrate geodata from several different mediators (*Figure 4*). This metadata and top-level ontologies can be maintained by shared GIS server (Catalog) as suggested by GeoNis architecture.

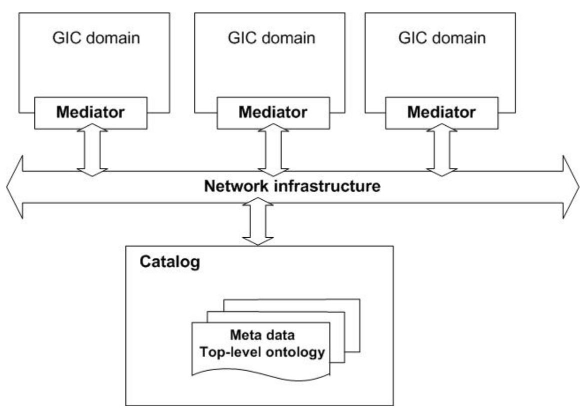

Figure 4*.* Semantic mediators chaining

SEMANTIC ONTOLOGY MAPPINGS

In ORHIDEA 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.

An ORHIDEA Semantic mediator uses hybrid ontology approach (Figure 5). Meaning of the terminology of each community is specified in the local ontology. GeoNis with ORHIDEA 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).

Ontologies in ORHIDEA can be represented as directed graphs where nodes correspond to concepts and arcs correspond to roles and *isa* relationships. We have defined set of relations between concepts in ontology (see figure 3): R = {synonym, hypernym, hyponym, meronym, T}, where T is a set of "topological" relations (which represent topological relations between real world entities): T={arc-node, route, node-route, point-event}.

The global terminology, represented by top-level ontology, can be seen as the set of basic domain terms. Hybrid ontology approach in ORHIDEA implies three types of semantic mappings between concepts from two local ontologies:

- Direct semantic relationships between two ontologies (on figure5, mark as ①),
- Indirect semantic relationships across top-level ontology (②, on figure 5), and
- Semantic mappings across reference (common) model (③, on figure 5).

Formal definition of ORHIDEA ontologies and semantic inter-relationships (of type ① and type ③) is given in (Stoimenov, 2003, Stoimenov, 2005). In this paper we will shortly present semantic inter-correspondences which are important in semantic conflicts resolution.

The relationship between concepts of different information sources (between local ontologies) is the task of the semantic inter-correspondences. 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,
- 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 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.
- 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.

Also, we define Refers_to(RefClass a, OntologyConcept c) relationships between reference (common) model object classes and application ontology classes (concepts). Predicate Refers_to enables definition of semantic relevance SRelev(b,c) inter-correspondence between concepts from different ontologies (on fig. 3, type ③). 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.

However, it might be impossible to define direct mappings between concepts from distinct local ontologies. The reason is very simple – for effective semantic mapping, experts have to define semantic inter-relationship between two pairs of local ontologies. So, it is obvious that it becomes impossible to provide explicit mappings between the concepts from local ontologies.

The mapping between local ontologies and the top-level ontology can be based on generic relationships, such as "Subclass_Of" and "Same_Class_As", and on relationships that depend on the application domain. In our approach, this type of mappings can be based on defined "one-to-one" semantic interrelationship SEqu, SContain and SIntersect. In this case, SEqu is used as “Same_Class_As” type of relationship, and SContain and SIntersect as “Subclass_Of”.

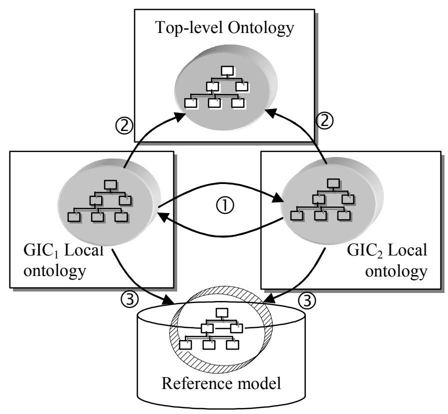

Figure 5: Hybrid ontology approach in GeoNis.

But, there are more complex relationships between general concepts form top-level ontology and concrete concepts from local ontology which we have to define. We have identified three more types of semantic inter-correspondences between TL and local ontology. The first one is “one-to-many” semantic mapping, where two or more classes (description of concepts) from local ontology together are a “Subclass_Of” one general class (description of concept) from top-level ontology. In that case, concept c1 from top-level ontology TLO is generalization of concepts C from local ontology O:

$$SHasParts(c_1,C) = \{(c_1,c_i) \mid \forall c_i,\ i=1,n \wedge c_1 \in TLO \wedge c_i \in O \wedge C=U\ c_i \wedge E(c_1) \subset U(E(c_i))\}$$

The second one is “many-to-one” semantic mapping, where one concept from local ontology is “Subclass_Of” two or more general concepts from TL ontology. This type of semantic relationship corresponds to multiple inheritance:

$$SMultiISA(C, c_1) = \{(c_i, c_1) \mid \forall c_i\ i=1,n \wedge c_1 \in O \wedge c_i \in TLO \wedge C=U\ c_i \wedge U(E(c_i)) \subset E(c_1)\}$$

Last one is “many-to-many” semantic mapping, where two or more concepts from local ontology together are “Subclass_Of” two or more general concepts from TL ontology:

$$SIntersectContent(C_1, C_2) = \{(c_i, c_j) \mid \forall c_i,c_j\ i=1,n,\ j=1,m \wedge c_i \in O \wedge c_j \in TLO \wedge C_1=U\ c_i \wedge C_2=U\ c_j \wedge U(E(c_i)) = U(E(c_j)) \}$$

This type of relationship is not useful for semantic conflict resolving, but if it exist, that means there is inconsistence in TL ontology. In this case we have to define new general concept in TL ontology as super class of M corresponding concepts. Also, we have to replace “many-to-many” relationship with new one “one-to-many” relation.

Inference module RuleEngine is important part of Semantic Mediator for resolving semantic conflicts. It provides inference services that allow more sophisticated queries to be formulated against the catalog feature objects and metadata. Knowledge in RuleEngine knowledge base is organized as IF-THEN rules. Such rules define semantic conflicts in form that is acceptable for inference process. For example, following rules defines some basic semantic inter-correspondences and basic relationships (c1 and c2 are concepts from different local ontologies, b is a concept from top-level or reference ontology):

R1: IF SContain(c_1,c_2) THEN TELL_KB(c_2)
R2: IF SEqu(c_1,c_2) THEN RETURN(c_2)
R3: IF Name(c_1) = Name(c_2) AND NOT_IN_KB(SNEqu(c_1,c_2)) THEN RETURN(c_2)
R4: IF SEqu(c_1,c_2) AND SEqu(c_2,c_3) THEN SEqu(c_1,c_3)
R5: IF SNEqu(c_1,c_2) THEN SNEqu(c_2,c_1)
R6: IF Refers_to(b, c_1) AND Refers_to(b, c_2) THEN SRelev(c_1,c_2)

Standard inference process implied only searching in ontology tree. The search can be performed by automatically mapping between concepts in the same ontology (within the same domain). This is possible by applying a standard terminological reasoner (for example, FaCT (Horrocks, 2003)), which can work with concepts described in the description logic.

Defined rules are used as part of inference process in case of searching connections between nodes in two (or more) ontology trees. Such connections are defined by semantic interrelationships and represented by corresponding IF-THEN rules. This procedure enables semantic query reformulation in a way that is not possible with mentioned standard search of ontology tree. Up to now, our knowledge base contains approximately 30 IF_THEN rules coded in description logic.

CONCLUSION

Semantic heterogeneity of the data sources in GeoNis is resolved by Semantic Mediator using a hybrid ontology approach. Meaning of the terminology of each community is specified in the local ontology. GeoNis 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.

This framework is comprehensive enough to manage various types of semantic conflicts in heterogeneous information sources while preserving the autonomy of individual sources. The principles behind the ontology/mediation framework are extensibility, relative autonomy of infrastructure nodes, and universal access to heterogeneous data sources from a variety of portals.

BIBLIOGRAPHY

Abel, D.J, Ooi, B.C., Tan, K.L., Tan, S.H., 1998. Towards integrated geographical information processing. Int. Journal of Geographical Information Science 12 (4), 353-371.

Bishr Y.A., Pundt H., Kuhn W., and Rdwan M., 1999, 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, 1999, pp. 55-70.

Egenhofer, M., 2002. Toward the Semantic Geospatial Web, Tenth ACM International Symposium on Advances in Geographic Information Systems. ACM Press, New York, NY, USA, McLean, Virginia, USA, pp. 1-4

Fonseca F.T., Egenhofer M.J., Agouris P., Camara G., 2002. Using Ontologies for Integrated Geographic Information Systems. Transaction in GIS 6(3), 231-257.

Genesereth M., Keller, A.M., Duschka, O.M., 1997. Infomaster: an information integration system, Proceedings ASIGMOD Conference, pp.539-542.

Hakimpour F., Timpf S., 2001. Using ontologies for resolution of semantic heterogeneity in GIS. Proceedings 4th AGILE Conference on Geographic Information Science, Brno, Czech Republic, pp.385-395.

Horrocks, I., 2003, The FaCT System, http://www.cs.man.ac.uk/~horrocks/FaCT/

IEEE 1990, IEEE Institute of Electrical and Electronics Engineers, 1990. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY.

Kuhn, W., 2002. Modeling the Semantics of Geographic Categories through Conceptual Integration. In: M.J. Egenhofer and D.M. Mark (Editors), GIScience 2002. Lecture Notes in Computer Science. Springer-Verlag

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

Meersman R., 1995. An essay on the role and evolution of data (base) semantics. In: Meersman R, Mark L. (Eds), Database Application Semantics, Proceedings of IFIP WG 2.6 Working Conference on Database Application Semantics.

OGC, 2004, OpenGIS Consortium Inc, OpenGIS Simple Features Specification For OLE/COM, 1999, http://www.opengis.org

Stoimenov L., Djordjevic-Kajan S., Stojanovic D., 2000. Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology. Proceedings MELECON 2000, 10th Mediterranean Electrotechnical Conference, Cyprus, Vol.1, pp.334-336.

Stoimenov L., Đorđević-Kajan S., 2002, "Framework for Semantic GIS Interoperability", FACTA Universitatis, Series Mathematics and Informatics, Vol.17 (2002), pp.107-125.

Stoimenov L., Đorđević-Kajan S., 2003, Realization of GIS Semantic Interoperability in Local Community Environment", Proceedings printed as book, Eds. M.Gould, R.Laurini, S.Coulderon, ISBN 2-88074-541-1, 2003, Presses Polytecniques et Universitares Romandes, 6th AGILE conference on Geographic Information Science, "The Science behind the Infrastructure", AGILE 2003, Lion, France, April 20-23.2003. pp.73-80

Stoimenov L., Stanimirović A., Đorđević-Kajan S., 2004a, "Realization of Component-Based GIS Application Framework", Proceedings printed as book, Eds. F.Toppen, P.Prastacos, 7th AGILE Conference on Geographic Information Science, AGILE 2004, Heraklion, Crete, Greece, April 29 – May 1, 2004., ISBN 960-524-176-5, 2004, Crete University Press pp.113-120.

Stoimenov L., Stanimirović A., Đorđević-Kajan S., 2004b, "Interoperable Component-based GIS Application Framework", Acta Electrotechnica et Informatica, ISSN 1335-8243, Izdavač: Faculty of Electrical Engineering and Informatics of Technical University of Košice, No. 4, Vol. 4, 2004, pp. 25-31.

Stoimenov L., Đorđević-Kajan S.,2005, An architecture for interoperable GIS use in a local community environment, Computers & Geosicence, Elsevier,.31 (2005), 211-220.

Stuckenschmidt H., Wache H., Vogele T., Vissar H., 2000. Enabling Techologies for Interoperability, In: Vissar U., Pundt H. (Eds.), Workshop on the 14th International Symposium of Computer Science for Environmental Protection, Bonn, Germany, pp.35-46.

Wiederhold, G. 1998. Weaving data into information. Database Programming & Design 11 (9), 22-29.