Abstract. Semantics of geospatial information is necessary to ensure its interoperability and correct usage. Geospatial ontologies enable encoding such knowledge and enable reasoning mechanisms to estimate interoperability. The Framework of interoperability for geospatial information using ontologies, presented in this paper, supports (i) building of ontologies, (ii) semi-automated, ontology-based mapping of geospatial databases and (iii) logic based elimination of inconsistent mappings (iv) consequent semantic translation of data across heterogeneous databases. We discuss the extension of this approach to facilitate interoperability of web feature services employed in spatial data infrastructures.

Keywords: interoperability, ontologies, geospatial-databases, web-services, semantic translation.

1 Introduction

Future scenarios for Geographic information systems could be plausibly imagined as a potential single infrastructure working together in harmony with a wide variety of geographic-data. This data, often collected independently, is usually heterogeneous in nature. Moreover the data that a given user requests may not be available from a single data source and it is necessary to assimilate data from more than one data source. As can be expected, such information retrieved from multiple sources would not be syntactically, structurally and semantically homogeneous and would be technically unusable. Manual integration usually employed in such cases can be very tedious and prone to error, especially at the time of emergencies when such integration is most required. Machine based integration has been evolved over recent years with the concept of ontology driven GIS (Fonseca and Egenhofer 1999) and ontology enabled databases (Sarda 2003)

Geographic Information Integration is the key to large scale and wide ranging use of geospatial data. Some of the heterogeneities that exist between different information systems can be classified as follows:

i) Data Heterogeneity:
   It refers to heterogeneity among existing systems in terms of data representation and storage etc and can be further classified as below (Bishr 1997).
   a) Syntactic Heterogeneity - in which, the information systems use different storage and representation paradigms.
   b) Schematic Heterogeneity - in which, the same object in the real world is represented using different concepts.
   c) Semantic Heterogeneity - in which a fact can have more than one description or interpretation.

ii) Functional Heterogeneity:
   This type refers to the heterogeneities among existing systems in terms of service interfaces.

Both types of heterogeneities result from the differences in (i) systems and technologies (ii) systems and standards (iii) semantics and domain specific knowledge. Such differences are abundant in most vertical sectors, which produce or use geospatial data. The typical usage of geospatial data includes different vertical and geographic domains. This aggravates the interoperability problems and has been reported by many researchers.

Interoperability itself comes in various stages and many classifications of such stages can be found in literature [1]. The classification of interoperability in terms of syntactic, structural and syntactic issues, in accordance with the heterogeneities discussed above, are important for the approach to resolve such issues. Open standards in geospatial data exchanges such as GML (OGC 2003a) help to resolve syntactic and structural differences. Many approaches to resolve semantic interoperability exist including that of ontologies and reasoning based on ontological specifications.

This paper presents work on the framework of semantic interoperability developed to deal with heterogeneous geospatial data. It uses, geospatial ontologies and mapping mechanisms between them to resolve semantic issues. The paper is arranged as follows: This chapter provides a background to the topic of semantic interoperability with an introduction to the specific problems of the Indian NSDI. Chapter 2 discusses the concept of ontology-enabled databases and the use of geospatial ontologies especially for large scale sharing of geodata, such as spatial data infrastructures. The subsequent section discusses the framework of interoperability based on (i) extraction (ii) management and (iii) mapping between ontologies. This framework also includes interfaces to automated semantic translation of data and data services based on the mappings created between given two ontologies. Finally in section 4 we discuss the results of our work and present some ideas for future work.

1.1 Background

Semantic heterogeneity of databases has received attention for many years now and ontology based mapping has been increasingly viewed as an engineering solution to the problems. Based on specifications of the conceptualizations (Gruber 1993) as a more generic layer above the schema specifications, ontologies serve as an intermediate step to specify and resolve semantics of the
contents of a database system (Fonseca, Egenhofer et al. 2002). Two categories of semantics can be differentiated in regard to

a. Classes or schema names and
b. Individuals or instances of the classes.

While the later is by no means a trivial problem, in this paper we restrict our discussion to the semantics at the schema level.

Although many definitions for ontologies exist, we know that they can be expressed in the form of hierarchies of interrelated synonyms, hypernyms and hyponyms of the vocabulary that defines the shared domain. This linguistic bias in the statement is motivated by the link of ontologies to computational lexicons such as WordNet (Fellbaum 1998). A tool for lexical similarity based ontology mapping has been developed and reported by Sen et al. (Sen, Somavarapu et al. 2006). Heuristics based algorithms are employed to propagate lexical similarity values from subclasses and attributes to reflect implicit meanings of class names based on their attributes and their position in the taxonomic hierarchy. Machine generated values of such similarities can be used to generate mappings across classes to enable semi-automated mappings.

Nevertheless there are several approaches to resolve semantic interoperability and the most common approach based on human intervention, is the most expensive and often incomplete. Issues such as omissions and errors of semantic understanding have acted as barriers for the large-scale use of geospatial data and any efforts to reduce such barriers can be considered as an advancement of the opportunities for greater and appropriate use of data. We now inspect the need for semantic interoperability to specify the exact requirements of a framework, which enables greater degree of semantic interoperability.

1.2 Requirements of semantic interoperability

It is necessary to consider the two main requirements to understand the need for semantic interoperability. The first is the need for quantification of existing of interoperability between two systems (so as to ascertain the existence or absence of semantic conflicts) and second, the need to ensure translatory mechanisms in cases where semantic equivalence can be established.

There are several attempts to achieve the two objectives in relation to semantic interoperability. Quantification of semantic inconsistencies between two systems based on the classification of the types of such inconsistencies is important but not always achievable nor comprehensive. Translationary mechanisms, on the other hand, have also been attempted with varied scales and successes. Mapping tools such [Mitre reference] and Rondo (Melnik, Rahm et al. 2003) are such developments and use ontologies or model management as the basis of such translatory mechanisms.

While exact quantification of semantic interoperability is almost impossible given the variable nature of semantics in relation to time, regions and other factors, it is important to note that several attempts have been made to describe naming heterogeneities and interoperability across systems (Fonseca 2006).

Geospatial ontologies are important tools, which help to achieve both above objectives and have been the cornerstone for semantic interoperability efforts (Chandrasekaran, Josephson et al. 1999). (Kuhn 2005). To understand the scope and limitations of such efforts we outline the concept of ontology enabled database systems, which introduces ontologies as a level above the structural layer of databases (Sarda 2003). We also discuss the special case of geospatial ontologies, which are being increasingly used in the geospatial domain.

2 Ontology enabled database systems

The term ontology enabled database systems (Sarda 2003) is used to represent the use of ontologies as intermediate layer above the schematic layer in the context of user-database interactions. As shown in the figure below, this includes several tasks, including

(i) Construction of ontologies, which specify the concepts, expressed in the database schemas
(ii) Storage and maintenance (including version control) of such ontologies
(iii) Mappings of concepts between different databases on the basis of their ontological specifications
(iv) Linkage to database query modules to ensure semantic translations of the query and the resulting datasets on the basis of the mappings.

Such an architecture is motivated by the notion of generic schema management (Alagic and Bernstein 2000) and draws on conventional database concepts combined with the understanding that database schemas are only a symbolic representation of the concept represented in the database. Hints about the nature of the conceptual entity are usually captured in the metadata and is often left to human users. Entity relationship diagrams and object oriented design diagrams (UML diagrams included) are informal specifications of such knowledge as well. Ontologies enable formal specification of such knowledge and also allow machine based reasoning. Such mappings are the basis of semantic translations, which are the core mechanisms of the OEDBMs.

It is important to note that geospatial databases used by different domains may have different class names and attribute names to denote the same entity or in the worse case, could use the same name to denote different entities1. Thus ontologies are essential to extend the notion of OEDBMs to the geospatial domain. Ontologies of the geospatial domain ontologies or geospatial ontologies are needed to develop such an upper layer of ontologies to resolve semantic conflicts in geospatial databases.

1 In the case where same class-names denote different concepts, it is often the issue that lexical similarity leads to errors. Such errors are also high in human based mappings as well and hence we say this is a more serious problem.
2.1 Geospatial ontologies

Ontologies of the geospatial domain define geographic objects, fields, spatial relations, processes and their categories. The view advocated by Câmara et al (Camara, Monteiro et al. 2000) of geographical space as a system of entities and a system of actions and hence the specification of geospatial ontologies as those concerning the specification of concepts of geospatial entities and geospatial actions. It has been shown by Sen (Sen 2007 (forthcoming)) that geospatial ontologies based on specifications of both the types of concepts is essential but in this paper we shall restrict our discussion to concepts of entities alone.

Geospatial knowledge sharing with ontologies aims to provide methodologies and algorithms to translate and map between the models of geospatial data and applications. They aim to provide the semantic knowledge for machine-based integration of geospatial information. Attempts to employ geospatial ontologies have been attempted from by many researchers and many approaches to extraction, use and representation of geospatial ontologies have evolved (Agarwal 2005). However, with the increased use of web based services of geospatial data and the emergence of tools in the semantic web, there has been increased focus on specification of geospatial ontologies and to study the shortcomings of present day approaches.

Geospatial ontologies, in the minimal specifications consist of concept names and their relationships. The primitive relation of subsumption (analogous, although not completely, to the is-a relation in object oriented designs) allows us to specify Directed Acyclic Graphs (DAGs) to represent the knowledge about certain geospatial entities. Such tree diagrams (as shown in Figure 2 below can be integrated with the knowledge about the attributes (or properties) of such entity classes. Specifications of such class diagrams along with properties are the basis of commonly found ontologies in the geospatial domain. Further specifications of axioms related to these concepts and attributes of these concepts are also important parts of geospatial ontologies but can be specified at later stages when such information is available.

Web ontology standards such as OWL (Antoniou and Van Harmelen 2003) and ontology engineering standards such as ontoclean (Guarino and Welty 2004) play an important role in the absence of geospatial ontology specification standards and processes.

![Figure 1](image1.png)

**Figure 1** Database management system extended for Ontology management. OEDBMS enables ontology definitions, browsing, querying and mappings.(Sarda 2003)

![Figure 2](image2.png)

**Figure 2** Hierarchy of road network concepts with attributes (outside the boxes). The arrow denotes parent to child relations.

2.2 Spatial data infrastructures and geospatial ontologies

As far as Geographic Information (GI) and Spatial Data Infrastructures (SDI) are concerned, the research community is aware of the potential benefits of using ontologies as a knowledge representation mechanism. For instance, [Nogueras-Iso et al, 2005] identifies three main areas for the application of ontologies in an SDI. First of all, they can be used for data sharing and systems development. Ontologies help to define the meaning of features contained in geo-spatial data and they can provide a "common basis" for semantic mapping, e.g. to find the similarity between two features that represent the same object but have been defined using different languages. In the same direction, ISO/TC211 proposed the standardization item 19126 [ISO, 2004] to create a data dictionary defining the features and attributes that may be of use to the wider international community.

Geospatial ontologies facilitate the classification of resources and information retrieval. Metadata ("data about data") describe unambiguously information
resources to enhance information retrieval, but this improvement depends greatly on the quality of metadata content.

Ontology based discovery of web services and chaining of services based on such discovery has been suggested (Lutz, Einspanier et al. 2004), (Bernard, Einspanier et al. 2003). The primary objective of such a mechanism is to ensure the facilitation of services using geospatial data to publish-bind-find services, which provide geospatial data. A typical example of such integration is shown in Figure 3 below. It is important to provide frameworks under which such chaining of geospatial web services is enabled in a spatial data infrastructure. We discuss such a Framework for Interoperability of Geospatial data using Ontologies (FIGO) in the next section.

**Figure 3** Service integration for Geospatial data. This example shows a typical service integration where a WFS serving road network data is chained to two routing services.

### 3 FIGO

Schema matching based on ontologies has been proposed in FIGO the main aspects of this framework include:

1. A workflow for semi-automated ontology based mapping of schema elements, which mainly relies on the schema level information.
2. Semi-automation is achieved based on machine-generated confidence values of similarity of attributes, classes and layers related to the schema of a particular source and that of the user. Heuristics based similarity value propagation is also employed to improve the confidence values.
3. Use of the ontology mappings to generate schema translations from the source to the target. This framework as shown in figure 1 includes both (i) mapping validation and (ii) Xquery wrapper generation, which form the components of the schema translation.

FIGO is developed on the basis of lexical similarities of class names and their attributes. This framework uses specifications of database ontologies based on a graph of layers, classes and their attributes. Layers are basically aggregates of certain classes grouped as per their usage. Both Layers and Classes can have child elements in the form of Sub-layers and Sub-classes respectively. Attributes can be both XML data-types and also object-types (contains Classes). This graph structure without further axioms about the classes and attributes themselves forms the basis of our simplistic ontologies for a given database as shown in Figure 4 below.

**Figure 4** Schematic representation of relations between Layers, Classes and Properties within FIGO.

#### 3.1 Ontology creation and management

Ontologies can be created within FIGO using a graphic user interface but they can also be imported from schema definition files. This process involves careful selection of metadata descriptions of the schema elements that are usually helpful in lexical mapping across ontologies.

Ontology management includes several versions of ontologies for a given data source (and user ontologies as well). These are also linked to the physical database structure so that the semantics specified are truly reflected in the database.

Creation of ontologies is usually a massive task and requires considerable help in terms of a shared vocabulary or definitions such as GML feature types and XML datatypes. Such basic definitions are available to users along with a facility to export the entire ontology into a text file, which could be later imported for yet another ontology.

Users can browse ontologies using a tree structure and define layers, classes and attributes as already shown in figure 1.

#### 3.2 Ontology Mapping

Ontology mapping is a semi-automated process within FIGO and uses lexical similarity along with similarity...
propagation algorithms to suggest confidence values for any given mapping between two ontologies. To understand the mapping between ontologies $O_S$ and $O_T$ we need to determine a mapping between their nodes.

The ontology-matching problem now transforms into generating a similarity matrix given by

$$M[O_S X O_T] = \{m_{S1T1}, m_{S1T2}, \ldots, m_{S1Tn}, m_{S2T1}, m_{S2T2}, \ldots, m_{S2Tn}, \ldots, m_{SmT1}, m_{SmT2}, \ldots, m_{SmTn}\}$$

such that $0 \leq m_{ST} \leq 1$

The matrix is directed from a given source to a target and the reversibility of the values is not assumed (i.e. the similarity is not symmetrical). The question that such a mapping answers is ‘which are the classes and attributes of the source ontology that can provide corresponding values for the classes and attributes of the target’. This is similar to the prevalent understanding of ontology mapping in literature. Ehrig and Staab (Ehrig and Sure 2004) define ontology mapping: “Given two ontologies $O_1$ and $O_2$, mapping one ontology onto another means that for each entity (concept $C$, relation $R$, or instance $I$) in ontology $O_1$, we try to find a corresponding entity, which has the same intended meaning, in ontology $O_2$.” The mapping assumes that inability of assessing the similarity results in a zero value, and hence such cases are treated as non-equivalent. The other relations used in the mappings are equivalent, sub-class, super-class and attribute union. Mappings with values lower than a given threshold ($t$) is assumed as non-equivalent.

The machine based mappings values serve as guidelines for manual mappings of the ontologies. These mappings can be represented as many to one correspondences directed from the source to the target. Attributes of the target could be mapped as computation (such as addition or string concatenation) of more than one attribute of the source. The mappings are not necessarily logically correct and can have in logical inconsistencies.

We apply logical reasoning upon such mappings to obtain a subset of consistent mappings, which are presented to a human user who confirms each correct mapping between two concepts of the two ontologies.

3.3 Eliminating inconsistent mappings

We adopt an approach to eliminate logically inconsistent mappings based on the hierarchical structures of the given ontologies. If class A is mapped to class B of another ontology, then inconsistencies are raised if (i) parent classes of A are mapped to child classes of B and in general (ii) when child class of A is mapped to a class which is not a child of B. This is shown in Figure 6 as below. Eliminating such inconsistencies gives major savings in labor costs for the human mapping exercise. We use automated sequential elimination of inconsistencies based on ontological constraints as an advancement of semi-automated mappings across domains.

After mappings have been validated and confirmed by the user, these are stored in the databases and can be used for query translations.
3.4 Query translations using ontology mappings

Now that validated mappings between ontologies of two given sources (or a source and an user) exists, it is now possible to translate the query of a given user in terms of the ontology of the source. A similar mechanism for the translation of the dataset returned using Xquery statements. We explain this process of translations based on the algorithm outlined as below.

Algorithm 3 Transform Data Request

\[ S_{\text{req}} \leftarrow \text{list of sources with which there exists a ontology mapping defined with the user ontology.} \]
\[ \text{RequestedFeatures} \leftarrow \text{list of features in the data request} \]
for all \( S_{\text{req}} \) in \( S_{\text{req}} \) do
  RequestForSource \( \leftarrow \phi \)
  for all \( \text{RequestedFeatures} \) in \( \text{RequestedFeatures} \) do
    if There exists a mapping for \( \text{RequestedFeatures} \) in \( S_{\text{req}} \) then
      \( F_{\text{req}} \leftarrow \text{mappings for} \text{RequestedFeatures} \) in \( S_{\text{req}} \)
      Add \( F_{\text{req}} \) to RequestForSource
    end if.
  end for
  Generate Request for the source \( S_{\text{req}} \) with the list of features in RequestedFeatures
end for.

The translated views of a given road network data based on a given user ontology is shown alongside with the original data in figure 7 as below.

4 Extending FIGO to web service interfaces

Service oriented architectures are increasingly replacing conventional client server and media based delivery of geospatial data. Spatial data infrastructures around the globe such as the NSDI India are investing in Service oriented Architectures (SoA) for the dissemination and distribution of geospatial data for the wide array of applications they are involved in(Harrison 2002). Open standards of such service based distributed geospatial data and application architectures enable wider use as compared to traditional architectures that have been used in the geospatial domain. These ensure syntactic interoperability by ensuring XML based data interoperability standards. Web Feature Services (WFS) and Web Mapping Services (WMS) are based on such open standards (OGC 2002; OGC 2003b) and many tools conforming to such standards are available as commercial and open source software.

However the question of semantics remains largely unsolved, although there have been many attempts in this area. The use of ontologies for semantics based service matching has been attempted (Lutz, Riedemann et al. 2003; Klien, Lutz et al. 2004), but in the absence of a shared vocabulary, it is extremely difficult to map geospatial concepts across geospatial data providers. We now discuss the utility of FIGO in the integration of Web based services of geospatial services, especially that of WFS.

![Figure 8 Integrating WFSs in FIGO](image)

The main steps involved to integrate a given web service in the form of a WFS to FIGO, as shown in the figure 8 above include the following:

1. Request all available features in different sources (WFSs) by getcapability(), then it gets stored in FIGO tool.

![Figure 7 Results of query translations in FIGO](image)
2. Receive the feature lists in an XML format.
3. Request `describeFeature()` for every feature in the featurelist.
4. Receive the XSDs for every feature (along with attribute names).
5. Include the XSD details to build the ontology of each source dataset.
6. Store the source ontologies.

Once this has been achieved we can now interface user’s ontology with FIGO. The different source ontologies in FIGO are mapped onto this user’s ontology. We also employ this user’s ontology to create a generic user page\(^1\). This step generates an XQuery statement in terms of the user’s ontology.

As shown in figure 9 below, the overall architecture includes the use of the generic query page to formulate the WFS requests based on mappings of user ontologies with sources. The WFS requests are then executed to obtain GML data. XQuery based wrappers are applied on the data to translate them as per the user’s ontology.

**Figure 9** Stepwise procedure to integrate multiple WFSs with FIGO and enable semantically heterogeneous queries by users.

It should be noted that the ontological structures available from the schema definition (XSD) provided by WFS is not rich enough for ontological modeling. It often requires further investigation and is more accurate in cases where the ontologies of the sources are independently availed from the ER diagrams of the databases serving the WFS in the background.

The algorithms employed to carry out the two tasks are outlined as below. It is important to note the two-step approach used. The first step is offline and includes the computationally intensive process of mapping concepts across ontologies. This process is often (at least in FIGO) with human intervention. However the second phase in which user formulates the query based on his/her own terms, is at real-time. The response based on (i) generating appropriate WFS requests and (ii) conversion of the received GML files based on the user’s ontology, is at real-time.

5 Conclusions and Future Work

We have discussed and shown the framework for interoperability based on geospatial ontologies. This framework provides a mechanism to introduce translationary processes based on ontology mappings. The ontology mappings themselves are semi-automated and based on lexical and structural similarity of the schema terms. The extension of FIGO for translation of data from WFS has also been discussed. It is important to note that although contents in the form of geospatial ontologies are necessary, it is the framework, which is important in large data sharing mechanisms such as spatial data infrastructures.

While the work reported in this paper is the first step towards establishing this framework there are several directions for future work in this area. These include

1. Specify Ontologies: Agencies, Data Providers, Application providers, Users and even GIS vendors need to specify ontologies of terms used. In the context of databases it is not only important to mention class names and their attributes but also semantic descriptions.
2. Employ Natural Language Processing techniques: Natural language descriptions are usually recorded for schema descriptions. It is important to utilize such information for similarity mappings. Choice of the natural language is also important such that adequate tools are available.
3. Share and learn: While it is not important for the user to think in terms of the ontology of a data source, it is important to learn the technique of identification and specification of mappings. Sharing of such mappings is expected to help the generation of mappings of other users.
4. Improve the real-time performance of FIGO in the translation of WFS data. This can be achieved by query optimisation and simplification.

Acknowledgments.

This work was supported by Dept of Science and Technology (DST, Govt of India). We also acknowledge support from Ordnance Survey, UK for permission to use their data in our experiments.

---

\(^1\) The concept of a generic query page is based on the construction of an XQuery based on class names, attribute names chosen from the user’s ontology along with standard query operations.
References


Fonseca, F. and M. Egenhofer, C. Davis and G. Câmara, (1999). Ontology-driven geographic information systems. 7th ACM Symposium on Advances in Geographic Information Systems, Kansas, MO.


Sen, S., S. Somavarapu, et al. (2006). Class structures and Lexical similarities of class names for ontology matching. ODBIS (VLDB Workshop on Ontologies-based techniques for DataBases and Information Systems Co-located with VLDB 2006), Seoul, South Korea.