Discovering Mappings between Ontologies in Semantic Integration Process

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

The use of ontologies in semantic translation is a viable approach to overcome the problem of semantic heterogeneity. In our framework (called GeoNis) for GIS interoperability, based on Semantic Mediator, semantic heterogeneity is resolved by using a hybrid ontology approach. This paper presents our approach to mappings between community terminologies (local ontologies) and mappings between local ontology and top-level ontology. As a main contribution, paper defines off-line discovering of mappings between concepts from two local ontologies.

KEYWORDS: *ontology, semantic heterogeneity, semantic mappings, GIS interoperability*

INTRODUCTION

In recent years, a large number of diverse, distributed and heterogeneous information sources (databases, knowledge bases, collections of documents, etc), become available over the Internet. GIS applications often have to process geo-data obtained from various geo-information communities. 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; Stoimenov, 2004). 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.

Today, semantics are a key factor of successful interoperability between GI systems. 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 paper is structured as follows. Section 2 describes related work on semantic interoperability of GIS and role of ontologies. In Section 3 we discuss how our framework, based on semantic mediators, works with semantic conflicts resolution problem. We also discuss types of ontology mappings between concepts from local ontologies and concepts from global terminology (top-level ontology). In section 4 of this paper we focus on our approach for off-line discovering and interpretation of mappings between two local/domain (GIC) geo-ontologies. The result is extension of knowledge base with new semantic mappings rules and given probability.

RELATED WORK

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). A semantic conflict, or heterogeneity, occurs when the same real world entity, modeled by two or more people, probably will not have the same modeling, even though it is representing the same phenomenon of the application's domain.

The use of ontologies as semantic translators is a viable approach to overcome the problem of semantic heterogeneity (Hakimpour, 2001; Fonseca, 2002; Stoimenov, 2003). Ontologies provide machine-readable semantics of information sources that can be communicated between applications and humans, the goal highlighted in any GI interoperability research. 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.

Semantic similarity measures play an important role in information retrieval and information integration (Rodriguez, 2003). In order to establish a correspondence among different representations of a same real world concept, that were defined in different schemas (or ontologies), it is necessary to recognize the common concept through the identification of similarities as well as conflicts among those schemas (Hess, 2004). Once ontologies have been integrated, similarity measures are applied to compare concepts. A recent work presents different measures for comparing concepts whose formal definition support inferences of subsumption, and local concepts in different ontologies inherit their definitional structures from concepts in a shared ontology (Rodriguiez, 2003).

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.

SEMANTIC ONTOLOGY MAPPINGS 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). Semantic interoperability in GeoNis, resolved by Semantic Mediator (Stoimenov, 2004), is the ability of sharing geospatial information at the application level, without knowing or, understanding terminology of other systems.

In GeoNis Semantic Mediator we propose a semantic based integration approach that uses multiple ontologies, instead of an integrated view (Stoimenov, 2005a). Our Semantic Mediator uses hybrid ontology approach. Our solution is to formally specify the meaning of the terminology of each GIC (i.e. local service or office) using local ontologies and to define a translation between each GIC terminology (local ontology) and shared domain terminology (in top-level ontology). In this context, ontologies are virtually linked by inter-ontology relationships, which are then used to indirectly support query processing. Semantic Mediator provides a methodology and software support for semantic mismatches (conflicts) resolving between terminologies. This methodology uses the defined ontology mappings between each community terminologies and a top-level ontology or the common data model (reference ontology).

We consider ontologies with inheritance relations *is-a* and typed roles between concepts. In an (local or top-level) ontology, inheritance relationships define a partial order over concepts and carry subset semantics. The following short definition describes ontologies as it is used in our scenario. GeoNis Ontology is an abstraction of domain of interest D, represented by triple *O=(C, R, isa)*, where $C= \{c_i \mid i=1, n\}$ is a set of concepts, $R= \{r_i \mid i=1, n\}$, is a set of binary typed roles (or relations) between concepts, and isa is a set of inheritance relationships defined between concepts. Set of semantic relations between concepts defines semantics of concepts and their relevance. We have defined the following set of relations between concepts in ontology: *R = {synonym, hypernym, hyponym, meronym, T}*, where T is a set of "topological" relations: *T={arc-node, route, node-route, point-event}*. More details about formal definition of GeoNis ontologies are given in (Stoimenov, 2003) and (Stoimenov, 2005b).

The relationship between concepts of different information sources (between local ontologies) is the task of the semantic inter-correspondences (Stoimenov, 2003). We divide the semantic conflict (semantic inter-correspondences) into four types (type 1): Semantic equality (similarity), *SEqu(c_1,c_2)*, Semantic dissimilarity, *SNEqu(c_1,c_2)*, Semantic intersection, *SIntersec(c_1,c_2)*, and Semantic contain, *SContain(c_1,c_2)*. Resolving semantic conflicts among ontologies in GeoNis is based on semantic mappings using those inter-correspondences (or semantic inter-relationships). Hybrid ontology approach in GeoNis implies three types of semantic mappings between concepts from two local ontologies:

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

The mapping between local ontologies and the top-level ontology (type 2) 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".

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

Standard inference process defined in GeoNis approach implied searching in ontology tree and semantic relationships between ontologies. The search can be performed by automatically mapping between concepts in the same ontology (within the same domain) and between two ontologies (two domains) using direct semantic relationships. This is possible by applying a standard terminological reasoner, which can work with concepts described in the description logic.

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. Also, there are more complex relationships between general concepts from 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 top-level and local ontology (Stoimenov, 2005b). We use proposed semantic relationships between local ontologies across top-level ontology for generation of IF-THEN rules, but with user assistance and intervention (on-line user intervention). Inference module RuleEngine of Semantic Mediator uses those rules as additional knowledge for resolving semantic conflicts. Such mapping is required when one system sends query to another GIC, every single time. Because of that, this approach is very expensive and it demands involvement of domain experts.

DISCOVERING MAPPINGS

When one system request data from another, we need more accuracy from semantic mapping process to minimize the errors and misspellings. The more Geographic Information Communities (GIC) gets involved, more local ontologies are created. Thus, mapping of ontologies becomes a core question. The problem of missing semantic inter-correspondences can lead to interpretation conflicts during interchanging of geo-data. These conflicts often can be resolved by human intervention, as in proposed GeoNis (on-line) approach of generation of IF-THEN rules. But, there is lot of problems

in knowledge acquisition process and process of construction local ontologies and in defining semantic mapping rules. In such process we need additional involvement and commitment of experts from domain of interests. Nevertheless, efficiency of that process is not satisfactory enough. Correctness of semantic mapping, if there does not exist direct semantic mapping, is about 30%. As a consequence, automatic or at least semi-automatic techniques have to be developed to reduce the burden of manual creation and maintenance of mappings. This process of interpretation conflicts solving can be automated only if there exits semantic mapping for any pair of ontologies.

Because of that, we try to solve this problem by introducing semantic mappings across global ontologies (top-level and common model ontologies), and by generating the rules (based on semantic similarity), which are used to expand knowledge about mappings. We try to develop an off-line process of discovering semantic relationships between concepts from different ontologies. This is an on-request (off-line) search of indirect semantic mapping and generating of new rules. The main task of off-line discovering process is to recognize possible semantic similarity between concepts from two ontologies.

In proposed approach we do not use concept attributes to check similarity. Our approach implies use of semantic relationship for defining probability for similarity of two concepts. From our point of view we want to develop a process of semantic discovering of existing semantic relationship and compare those relationships between two concepts to find similarity among them. We use a definition of similarity as predicate $Sim(c_1,c_2)$, with meaning "probability of similarity between concepts c_1 and c_2", where c_1 and c_2 are the concepts from two different local ontologies. This predicate has values from: 0 (with meaning: false, or two concepts are dissimilar) to 1 (with meaning: true or two concepts are semantically similar). Predicate $Sim(c_1, c_2) = 1$ (two concepts c_1 and c_2 are similar), if there exist relationship r between c_1 and c_2, $r(c_1, c_2)$, and $r \in R$ (see definition of GeoNis ontology $O=(C, R, isa)$). Predicate $Sim(c_1, c_2) = 0$, (this is start value in process of discovering mappings), when two objects are different they don't have any kind of inter-relationships. Other values between 0 and 1 ($0<$ and <1), are the probability for semantic similarity of two concepts. Those values are calculated in process of discovering mappings.

What is the meaning of probability for similarity in the context of ontologies? The basic assumption is that knowledge about semantic is captured in semantic relationships within ontology. Based on the consistent semantics defined in ontology relations, it is possible to derive additional knowledge such as, in our case, similarity of concepts in different ontologies.

Calculation of the function $Sim(c_i,c_j)$ is the main task of our algorithm for discovering mappings. We compute probability for similarity between two concepts (from different ontologies) using existing semantic inter-correspondences with concepts from global ontologies (top-level and reference model), and also, by using inner-relationships between concepts (for now, only generalization/specialization *is-a*, and *synonym* relations) in a global ontology. When probability (value of $Sim(c_i,c_j)$) is grater than given threshold, our algorithm introduces new semantic similarity inter-relationship in knowledge base.

Algorithm for discovering mappings is following:

   1. Given two ontologies A and B, and two top level ontologies Gi, (i => global terminology/ontology or reference model), discovering mapping starts from one concept c_A from ontology A.

   2. For each pair of concepts from A and B, it is defined start probability for similarity (usually as 0). Repeat the following steps for each pairs of concepts c_A (from A) and c_B (from B).

3. If there is direct semantic inter-correspondence for concept c_A from A and concept c_B from B, skip it (semantic probability is set to 1).

4. Find a corresponding concept c_{G1} from top level ontology, which has the same or similar semantics (using semantic inter-correspondences) with concept c_A from ontology A (as illustrated on figure 1).
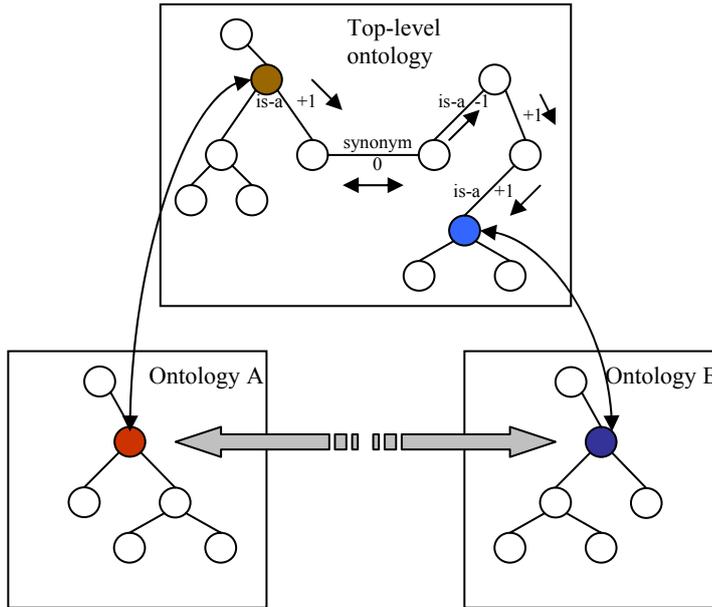

Figure 1. Discovering mappings between ontologies

5. Find a corresponding concept c_{G2} from top level ontology with same or similar semantic with concept c_B from ontology B using the same relations (as illustrated on figure 1).

6. Using defined set of inner-relations between concepts in top-level ontology, our algorithm tries to find connection between concepts c_{G1} and c_{G2}. This is practically graph path finding algorithm, where the length of path is calculated as number of *is-a* or *synonym* relations between c_{G1} and c_{G2} in the global ontology.

7. If there are more paths between two nodes/concepts c_{G1} and c_{G2} in top level ontology,

    7.1. For each path, compute value of probability predicate $Sim(c_{G1},c_{G2})$ as:

$$Sim(c_{G1},c_{G2}) = 1/(1+|Sum(p_i)|),$$

where p_i is weight of edges between vertex c_{G1} and c_{G2} in ontology graph (i.e. path length); as illustrated on figure 1, possible values of weight of edges are: +1, for $isa(c_{G1},c_{G2})$; -1 for $isa(c_{G2},c_{G1})$; and 0 for $synonym(c_{G1},c_{G2})$.

    7.2. Compute cumulative similarity for c_{G1} and c_{G2} as:

$$Sim(c_{G1},c_{G2}) = \max(Sim_{pi}(c_{G1},c_{G2}) \mid \text{for all paths } p_i \text{ between } c_{G1} \text{ and } c_{G2})$$

8. Go to step 4, and try to find similarity between another pair of concepts c_{Ai} and c_{Bi} from A and B.

9. If they exist, for each two pairs of concepts from G_i, compute cumulative probability as:

$$Sim(c_{Ai},c_{Bi})=Sim(c_{Gi},c_{Gj})+Sim(c_{Gk},c_{Gl}) - Sim(c_{Gi},c_{Gj})*Sim(c_{Gk},c_{Gl})$$

10. If $Sim(c_{Ai},c_{Bi}) > threshold$, Add semantic equality relation in KB with computed factor of probability $Sim(c_{Ai},c_{Bi})$ (concept c_{Ai} is probably mapped onto c_{Bi}, i.e. c_{Ai} and c_{Bi} are semantically similar with probability factor $Sim(c_{Ai},c_{Bi})$).

11. Go to step 2.

Our mapping approach is based on simple similarity measures. Our implemented approach is based on manually encoded rules for computing of probability. Those rules correspond to fuzzy logic rules for computing probability factor in a rule based system with inexact reasoning. The level of similarity, appropriate to indicate equality/similarity for the mapping is defined by the threshold. Every similarity value above the cut-off threshold indicates a match; everything below the cut-off is dismissed. For this method a fixed constant has to be defined as threshold, usually 0.8 or greater. All matches with a similarity above this constant are valid. Probability value for each path can be found through maximizing over different values.

CONCLUSION

The goal of this approach is to find the most correct mappings between two concepts from different ontologies. Our approach is based on three types of semantic mapping: semantic equal (similarity of concepts), semantic equal with given probability, and not semantic equal. Our approach of semantic mapping is extended with off-line discovering mappings. It is a three layer (semi)automatic approach having: direct mappings between two ontologies, mappings to be confirmed manually (by user intervention), and off-line discovering mappings.

The significance of our work is based on usefulness of GeoNis tools and components for realization of interoperable geo-spatial and other information nodes in local community organizations (such as B2B). A fully interoperable organization is able to maximize the value and reuse potential of information under its control. It is also able to exchange this information effectively with other equally interoperable bodies, allowing new knowledge to be generated from the identification of relationships between previously unrelated sets of data.

REFERENCES

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.

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.

Hess, G., Iochpe, C., 2004., Syntactic and Semantic GDB Conceptual Schemas Integration, in: 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.271-280.

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

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.

Rodriguez, M.A, Egenhofer M., Determining Semantic Similarity Among Entity Classes from Different Ontologies, IEEE Transaction on Knowledge and Data Engineering, 2003.

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., 2004, 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., Đorđević-Kajan S.,2005a, An architecture for interoperable GIS use in a local community environment, Computers & Geosicence, Elsevier,.31 (2005), 211-220.

Stoimenov L., Stanimirović A., Đorđević-Kajan S., 2005b, Semantic Interoperability using multiple ontologies, Proceedings printed as book, Eds. Fred Toppen, Marco Painho, AGILE 2005, 8th AGILE Conference on GIScience, Estoril, Portugal, 26-28.5.2005. , pp.261-270