Ontology Alignment for the Semantic Integration of Heterogeneous Geospatial Data Sets

Isabel F. Cruz
Department of Computer Science
University of Illinois at Chicago
ifc@cs.uic.edu

Position Statement

As a growing amount of networked information is becoming available, the potential of a large variety of applications accessing that data in a flexible and transparent way using a single query is far reaching. Particularly important is the case where hundreds of such data sources need to be queried in geospatial, scientific, educational, and e-commerce applications. For example, in homeland security applications, queries on a variety of different domains, e.g., geospatial, biological, and governmental, should be deployed with relatively little effort.

A fundamental issue is the integration of data whose schemas display semantic heterogeneities. We have been working on an approach where no *a priori* integration of all the data schemas involved is needed. Instead, querying can be easily extended to a new data source [3, 5]. Our approach has been used in real-world scenarios in the geospatial domain [1, 4, 6] and leverages current or emerging Semantic Web standards for the storage, interchange, and processing of data.

We have adopted an *ontology-based interoperation* mechanism, whereby a global ontology and the ontologies that describe the distributed data sources need to be *aligned* so as to establish *mappings* between the global ontology and each distributed ontology. A homogeneous view over all the data sources is thus obtained [2, 7]. We have investigated ways to help experts establish such mappings using a combination of manual methods, which are needed for the accuracy of the mappings, and automatic methods, which are needed to facilitate the experts' tasks. Those mappings produce *agreements*, which are declarative and therefore easy to understand and to maintain. Such agreements are automatically incorporated into the queries that result from the translation of the end user's query, which is expressed using concepts in the global ontology.

The *Agreement Maker* is a software tool that we have been designing, implementing, and user testing. It is used to create the mappings between the global ontology and each distributed ontology and to generate an agreement document. Figure 1 shows a user interface that is part of the Agreement Maker. The two hierarchies that represent the ontologies are displayed in two separate panes. The expert browses through the contents of the ontologies and establishes the mappings between each concept (or group of concepts) in the global ontology and the corresponding concept (or concepts) in the distributed ontology. A deduction module is integrated into the Agreement Maker to implement the automatic methods that infer new mappings, which are propagated along the ontotologies [7].

Open research issues include:

- Refinement of the mapping process, by exploring "similarity layers" between concepts and by defining new mapping types that can capture a wider range of semantic heterogeneities.
- Further automation of the mapping process by using other matching techniques, including similarity scores between concepts and multiple dictionaries.
- Experimentation with large ontologies to determine the efficiency of the automation process and its dependency on the topological properties of the graphs that are involved.

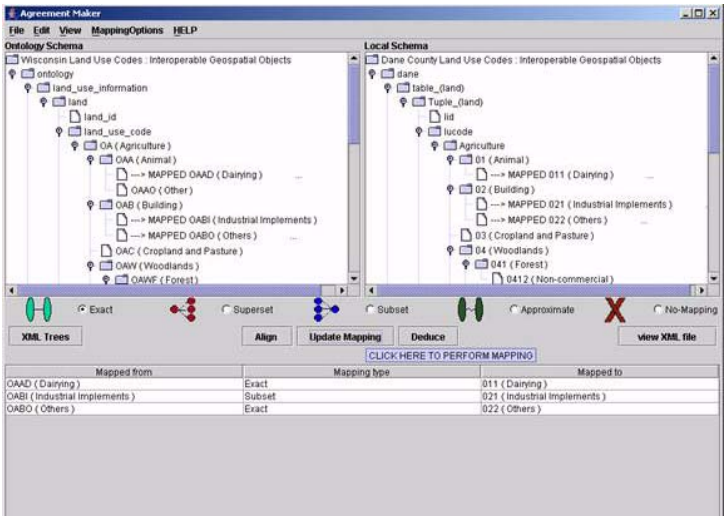

Figure 1: User interface showing established mappings.

- Refinement of the query translation process so as to allow for complex mappings to be incorporated into the queries.
- Extension of the current approach, which deals with a single theme, to data integration across several themes.

References

[1] I. F. Cruz and P. Calnan. Object interoperability for geospatial applications: A case study. In I. F. Cruz, S. Decker, J. Euzenat, and D. McGuinness, editors, *The Emerging Semantic Web*, pages 281–295. OIS Press, 2002.

[2] I. F. Cruz and A. Rajendran. Exploring a New Approach to the Alignment of Ontologies. In *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, in cooperation with the International Semantic Web Conference*, 2003.

[3] I. F. Cruz and A. Rajendran. Semantic Integration in Hierarchical Domains. *IEEE Intelligent Systems*, pages 66–73, 2003.

[4] I. F. Cruz, A. Rajendran, and W. Sunna. XML Database Integration for Visualizing US Election Results. In *National Conference on Digital Government Research*, pages 403–406, 2002.

[5] I. F. Cruz, A. Rajendran, W. Sunna, and N. Wiegand. Handling Semantic Heterogeneities Using Declarative Agreements. In *ACM GIS 2002*, pages 168–174, 2002.

[6] I. F. Cruz, W. Sunna, and A. Chaudhry. Ontology Alignment for Real-World Applications (Poster). In *Conference for Digital Government Research*, 2004.

[7] I. F. Cruz, W. Sunna, and A. Chaudhry. Semi-Automatic Ontology Alignment for Geospatial Data Integration. In *GIScience 2004*, LNCS. Springer Verlag, 2004.