Ontology-Based Geospatial XML Query System

Nancy Wiegand and Naijun Zhou
University of Wisconsin
550 Babcock Drive
Madison, WI 53706
wiegand@cs.wisc.edu, nzhou@wisc.edu

Isabel F. Cruz and William Sunna
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
ifc@cs.uic.edu, wsunna@cs.uic.edu

1. Introduction

This demo illustrates querying heterogeneous geospatial data over the Web. The Geospatial community is in the process of developing Web sites that enable coordinated access to geospatial data. A new example is the Federal government's Geospatial One-Stop portal [GOS03]. The purpose of this Federal initiative is to have all geospatial data accessible from one site. There are many other examples of geospatial Web sites including the Wisconsin Land Information System [WLISsite] and the Alexandria Digital Library project [ADL]. These Web sites allow keyword search of data sets but do not have a query capability for full data manipulation nor do they accommodate heterogeneous data.

To show the potential for adding a query component to these sites, we present our prototype system, which is modeled after Geospatial One-Stop (Figure 1). We add a link (*Query Data Content*) that launches our query management system, an XML-based Internet DBMS.

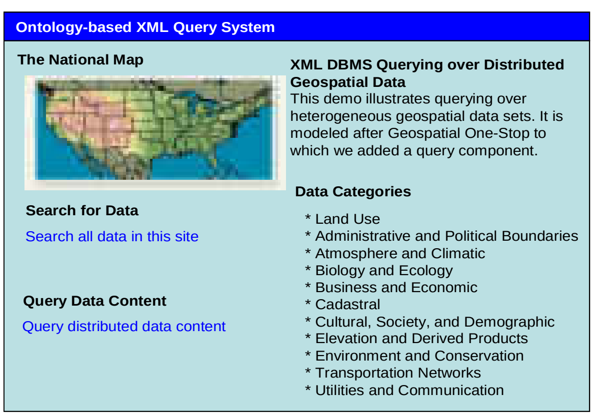

Figure 1. Demo Interface with a Query Link

2. Semantic Heterogeneity

In addition to providing a query component to a geospatial Web site, we address the problem of querying highly heterogeneous information that was developed by different government units and the private sector. As one moves across jurisdictional boundaries, database schemas vary, and the definitions and acceptable values for attributes change significantly. In fact, most of the effort required to integrate diverse geospatial data lies with the nonspatial attributes and not with geometric conversions.

Our system resolves semantic heterogeneity at the attribute and value levels for typical themes (data categories or channels) that are part of Geospatial One-Stop or the Wisconsin Land Information System (WLIS). Attribute level resolution is needed to resolve various identifiers for similar fields. Furthermore, value level resolution is needed to resolve complex heterogeneity problems in the domains of fields such as land use codes, which vary greatly between jurisdictions. For example, *cropland* in one jurisdiction may have to be compared to *cropland/pasture* or to *general agriculture* in other jurisdictions when determining agriculture usages over a multi-county area.

3. Ontology Mapping

Because such complex semantic comparisons between domain values of different data sets cannot be done using a simple algorithm or a fully automatic method, we use ontologies as a solution for semantic integration. An ontology is a shared and machine-executable conceptual model in a specific domain of interest [BFM02]. We express ontologies in XML. Because many geospatial data sets are in a relational table format, we use a table-based ontology DTD to describe schema, attribute, and value level information. For each attribute element, our ontology DTD includes an optional "**attribute value**" element to hold ontology *values* for heterogeneous domains. Furthermore, an attribute value element can itself have nested attribute values to be able to describe any level of subcategories for hierarchical domains, such as land use coding systems.

We developed a tool that uses a Local-As-View (LAV) approach to map a global ontology to each local schema. The tool first performs semi-automatic alignment over ontology trees using a deduction algorithm [CR03, CSC04]. Then, for mappings that cannot be resolved, a local domain expert uses the tool to choose a mapping between a global ontology value and a local value. The tool automatically generates an *agreement* file in XML for each source. The agreement file stores mapping types such as *subset* or *approximate* and is consulted by the query re-write system to generate subqueries.

4. Web Query System for Heterogeneous Geospatial Data

To provide full DBMS-type querying over distributed geospatial data, we use the Niagara Internet XML DBMS [NDM+01] as a base for our system. Niagara satisfies the need for general purpose querying over distributed XML data on the Web but does not have semantic integration facilities. We modified the Niagara Java source code to add an ontology subsystem. The ontology system is used to mediate between the user query and the local data. We also added metadata indexes containing minimal metadata needed to find data sets involved in a query and ontology indexes for agreement files.

We particularly address a typical type of query that occurs, for example, when land use managers pose a comprehensive query across multiple jurisdictions such as "*Find all cropland over a watershed that spans several counties*". Such a query requires mapping between heterogeneous coding systems. Our user interface captures a query containing the selection of multiple geographies and a predicate specifying a land use code. The predicate is expressed in ontology terms presented to the user. The type of query we address is different from a typical DBMS query because more than one data source is identified, but there is no join. We refer to this type of DBMS query, with the same predicate being applied to multiple geographies, as a G*eoQuery*.

To formalize the expression of this general query, we developed a *GeoSpace* concept, which is somewhat similar to the idea of an XML Namespace. A GeoSpace statement, added to XML-QL, contains a variable to hold the list of URLs for the data sources needed in the query. The variable is used in the body of the query as a qualifier for the generic ontology terms.

A domain interpreter consults the ontology mappings (agreement files) to resolve the general query to native terms for subqueries sent into Niagara. We enhance Niagara's output by providing semantic mapping information, additional summary statistics, and spatial displays. A full description of our system is given in [WZC+04].

Acknowledgement

This work was supported by the Digital Government Program of NSF, Grant No. 091489.

5. References

[ADL] Alexandria Digital Library project, http://webclient.alexandri.ucsb.edu.

[BFM02] Bussler, C.; Fensel, D.; and Maedche, A. 2002. "A Conceptual Architecture for Semantic Web Enabled Web Services", SIGMOD Record, Vol. 31, No. 4, pp. 24-29.

[CR03] Cruz, I.F. and Rajendran, A. 2003. "Exploring a New Approach to the Alignment of Ontologies", Semantic Web Technologies for Searching and Retrieving Scientific Data Workshop, International Semantic Web Conference, Sanibel Island, Florida, pp. 7-12.

[CSC04] Cruz, I.F.; Sunna, W.; and Chaudhry, A.; 2004. "Ontology Alignment for Real-World Applications", In Proceedings of The National Conference on Digital Government Research, dg.o2004.

[NDM+01] Naughton, J. ; DeWitt, D. ; Maier, D. ; and others. 2001. "The Niagara Internet Query System", IEEE Data Engineering Bulletin, Vol. 24, No. 2, pp. 27-33.

[GOS03] Geospatial One-Stop, http://www.geodata.gov.

[WLISsite] Wisconsin Land Information System Core Node Pilot Project, http://wlis.dnr.state.wi.us/wlis.

[WZC+04] Wiegand, N.; Zhou, N.; Cruz, I.; and Sunna, W. 2004. "A Web Query System for Heterogeneous Government Data", In Proceedings of The National Conference on Digital Government Research, dg.o2004.