Mingzhen Wei
Center of Excellence for Geographic Information Science
U. S. Geological Survey
mwei@usgs.gov

Enabling Semantic Power of Geographic Data

Geographic data are playing increasingly important roles in location-based applications, such as emergency rescue and travel planning. Geographic data storage is increasing at an exponential speed, and information and knowledge embedded are huge in amount and are extremely valuable for research and daily use. The purpose of this research is to enable semantic functionalities for the geospatial databases and to facilitate data access and integration in designing varied applications where geographic data are critical. My research and interest have been focused on two general aspects:

- Studying components required to enhance current geographic information systems (GIS) for enabling their semantic power, and investigating the enabling mechanisms.
- Studying ontology engineering approaches to assist ontology construction from available spatial data standards by the U.S. Geological Survey (USGS), Federal Geographic Data Committee (FGDC), and Open Geospatial Consortium (OGC).

GIS enhancement

Most current (2008) research on semantic enhancement to information systems (IS) is for relational database systems because of the power of the relational model in data management. The methods to add semantic models fall into two general categories. The first is to convert the entire dataset to Resource Description Framework (RDF) or one of its successors, then send semantic queries to generated semantic models to find the result; the second is to map relational datasets to semantic models, translate semantic queries to Structured Query Language (SQL), and pass the SQL to the original databases as in D2RQ [Bizer & Seaborne, 2004]. I am interested in the second option in my research because:

- Converting entire datasets to RDF results in large size of triple statements, creating a large number of triple management issues that were solved or are intensively studied in the database community. In addition, converting entire datasets to RDF wastes the power of original database management systems (DBMS) that have been well developed and well optimized in most cases (for example, Oracle, SQL Server, and ArcGIS database engine) for efficient data access and analysis. The original DBMS is indispensable because semantic models cannot do specialized data analysis without complicated remodeling.
- Database and semantic models describe the same objects in the real world in a different way. For example, database models take the closed world assumption and anything that is not included does not exist; whereas the semantic models take the open world assumption which makes the models flexible for update and enrichment. Regardless of the differences in their roots, two types of models have components that function similarly. Each type of model has advantages and disadvantages; therefore, it is of great importance to take the advantages of both types of models to improve data interoperability, accessibility, and real-time update.

The goal is to construct an enhanced GIS that can take advantage of both models and systems and enhance data searching and knowledge representation capabilities using semantic inference power. The proposed conceptual design includes the following components:

1. Construct ontological models from a given geospatial database that can be categorized as application-based ontologies. D2RQ represents a good study in this aspect, but it does not consider issues caused by database normalization and fine semantics hidden in the database. This research tries to generalize the rules for constructing ontologies and extracting fine semantic relations in a dataset.
2. Construct mapping between application-based ontologies and datasets. In considering data integration, formal definitions for a DBMS, a dataset, and other components in the dataset as Kashyap and Hanagan (2007) did for relational DBMS. With this mapping between them, the plan is to study the feasibility of exchanging some of functionalities in these two systems. For example, weather and how the data analysis functions in Oracle DBMS can be passed by semantic queries.
3. Translate semantic queries (for example SPARQL for RDF) to SQL for querying the databases. To take advantage of the semantic models and enable semantic reasoning, some inference rules may be required in the translation. For example, when a semantic query is sent to a subclass, corresponding SQL can be translated to query the large scale object and narrow down by defining filters, or to query directly the large scale object to eliminate the affect of misclassification for subtle subclasses. Although semantic and structured query languages are close to each other in structures, it is challenging to translate a semantic query to a SQL in an automatic and systematic way.
4. Translate SQL query results back to semantic representation so that it is possible to web-publish the result and to invoke another semantic query. By doing 3 and 4, a semantic cover [Malhotra, 2007] to the geospatial databases is constructed, and a dynamic semantic model is generated for instance data in result, assuming this semantic model for data representation and future analysis is needed. Based on domain and application ontologies and instance semantic models, knowledge can be inferred and stored in the knowledge base for future use.

Ontology engineering

Regarding domain ontology construction, great efforts and achievements have been made by the spatial communities including USGS, FGDC, and OGC. Those spatial data standards and metadata standards are valuable in constructing domain ontology for geographic data and features. As for duplicated efforts from different communities, I am interested in studying methods (for example formal concept analysis [Kavouras & Kokla, 2002]) for possible consensus for the following ontology integration: 1) Constructing domain ontologies from different spatial data standards for geographic data and geographic features; and 2) Integrating domain ontologies and database-based application ontologies.

References

Bizer, C., Seaborne, A., 2004, D2RQ — Treating non-RDF databases as virtual RDF graphs: Poster at 3rd International Semantic Web Conference, 2004.
Kashyap, V., Flanagan, M., 2007, From web 1.0 →3.0: is RDF access to RDB enough?: Position paper for the W3C Workshop on RDF Access to Relational Databases, Cambridge, 2007.
Kavouras, M., Kokla, M., 2002, A method for the formalization and integration of geographical categorizations: International Journal of geographical Information Science, 16(5), 439-453, 2002.
Malhotra, A., 2007, Integrating data from relational databases using RDF: position paper for the W3C Workshop on RDF Access to Relational Databases, Cambridge, 2007.