A Conceptual Design Towards Semantic Geospatial Data Access

Mingzhen Wei[1], Tian Zhao[2], Dalia Varanka[1], E. Lynn Usery[1]

[1]U.S. Geological Survey, Rolla, MO, 65401, USA, {mwei, dvaranka, usery}@usgs.gov

[2]University of Wisconsin-Milwaukee, tzhao@uwm.edu

ABSTRACT

This paper presents a conceptual design for connecting ontological and geospatial knowledge bases so that the system can take advantage of the semantic power of ontologies and data storage/analysis/query power of geospatial systems. The design is based on common-place queries, such as "what is a canyon/route/road" or "where is the canyon/route/road X."

Keywords: conceptual design, semantic data retrieval, geospatial database, ontology

INTRODUCTION

As the civilian national mapping agency of the United States, the U.S. Geological Survey (USGS) is developing *The National Map*. The goal is to provide consistent and high quality geospatial data to the public for decision making and knowledge sharing. Because the demand for data integration and sharing is increasing, *The National Map* adds semantics, intelligence, and user-centered design to enhance the characteristics in flexibility and user friendliness. An ontology is a formal and explicit specification of a conceptualization results for a domain of interest. Adding semantics to data modeling, ontology-driven systems, as a goal, will aid in data integration and data access, and facilitate different applications for *The National Map*.

An ontology is designed to model, reason, and infer feature types and their semantic relations. Geodatabases are designed to be efficient in storing large volumes of geospatial data, conducting geoprocessing, and responding to geospatial queries. In this paper, a conceptual design is presented to integrate ontology and geospatial database management

systems (DBMS) in a seamless way. The connection will enable the two systems to work at their best capacity, without having to generate redundancies.

PROBLEMS IN CURRENT GIS

In a geographic information system (GIS), geospatial and non-spatial data are stored in tables in relational database management systems (RDBMS). An example is shown in figure 1. To generate accurate results, a query must find the feature class represented as a table, or find the proper path to it, query the attributes (selecting the appropriate columns in a table), and construct filters with appropriate data values. Without assistance of well-designed interfaces, the complexity of GIS and geospatial data excludes casual users.

In constructing ontology-driven systems, the Web Ontology Language (OWL) and its subset the Resource Description Framework (RDF), are widely used for ontology definition. OWL organizes data as *classes, properties, and individuals*. The classes are sets of individuals and OWL has constructs to define set relations including subclass, equivalence, intersection, union, etc. In our setting, OWL *classes* would correspond to geospatial feature types, and OWL *datatype properties* to feature attributes such as "Road has name" (figure 2). OWL *object properties* describe relations between OWL individuals. The object prorperites such as "Road connectsAt Junction" may or may not have explicit counterparts in a geospatial database and may have to be derived. These constructs and semantic constraints enable OWL to specify complex semantics of geospatial data, and allow reasoners to check data consistency and to query individuals based on specifications of classes and properties. Although OWL provides powerful constructs for semantics, it does not provide direct translation from a database to an ontologic model, neither the mapping between their components. We need other tools such as Jena to build the mapping.

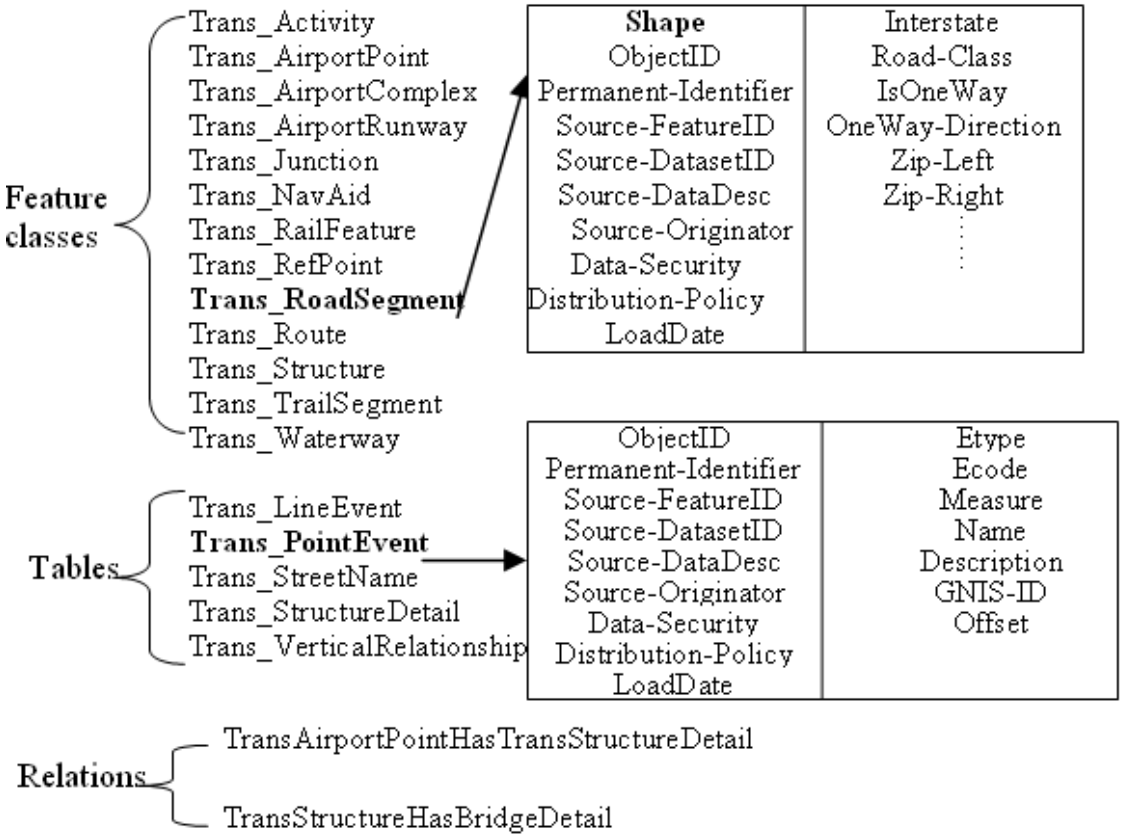

Fig. 1. Transportation data model in the Best Practice model for *The National Map*.

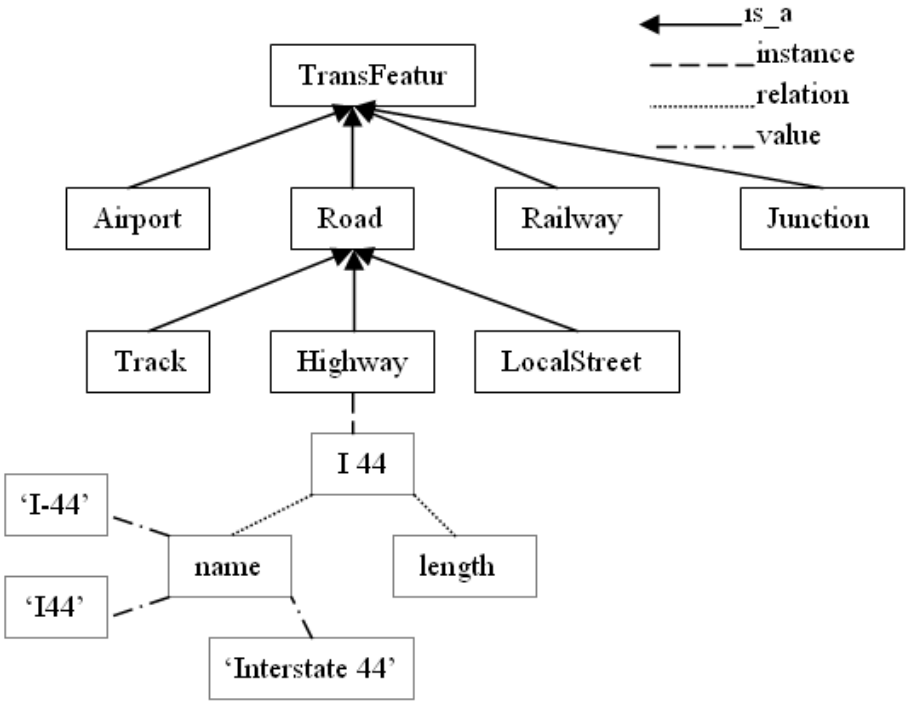

Fig 2. Data structure of ontology classes, relations, and values.

PRELIMINARY ARCHITECTURE

Geospatial ontologies and databases model data and knowledge in the geographic domain, and they have functionally comparable components (table 1); hence, a logical connection between them can be mutually beneficial. The integrated system will have the reasoning and inference ability based on ontology definitions; the ontology also can be constructed using the domain and feature knowledge from geo-databases, including direct and derived information. For example, some geodatabases save vertical relations of features, which can be expressed in ontological models as object properties “aboveFeature” and “belowFeature”. In addition, data in the geospatial database will have only virtual representation in the ontology and will not require legacy data replication.

Table 1. Data structure correspondence between ontologic and geospatial knowledge bases

Ontologic knowledge base	Geospatial knowledge base
Class	Table (e.g., feature class, domain)
Datatype property	Attribute, column in a table
Object property	Feature relation
Constraint	Field constraint, domain constraint
Individual	Instance, record

Ontology-driven systems have been proposed to aid intelligent data/knowledge management (Hammitt and Beckert, 2007). Sytems are proposed to connect spatial DBMS and ontologies (Dobear and Hart, 2006). Our preliminary design is illustrated in figure 3. The kernel includes three components: an ontology knowledge base (OKB), a geospatial knowledge base (GSKB), and a data exchange knowledge base (DEKB).

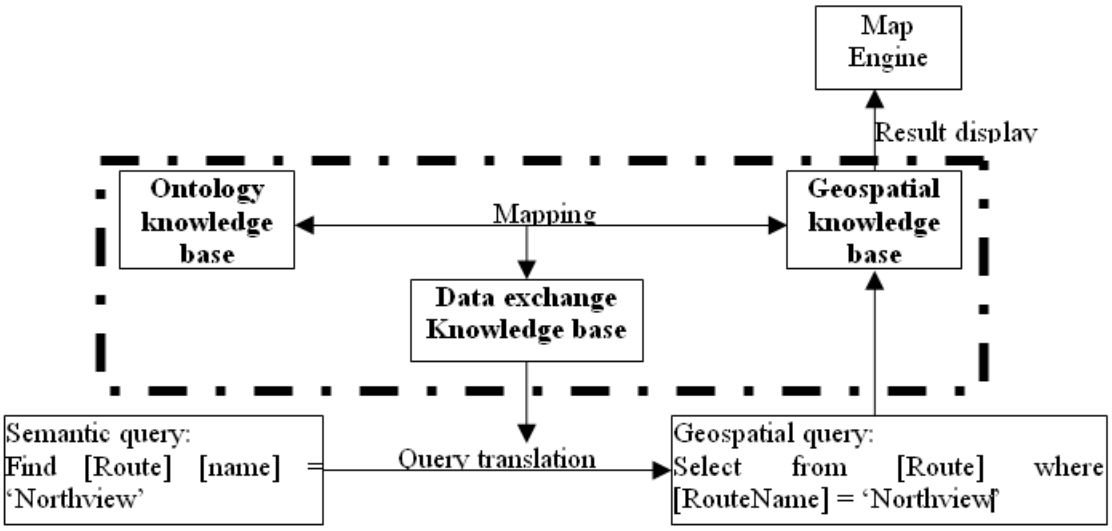

Fig. 3. The conceptual design toward semantically geospatial data querying.

An OKB includes ontology software environment, data storage mechanisms, geospatial feature models and instances, rule bases, knowledge from reasoning, and inference engines. A GSKB refers to GIS, the geospatial data, data structure, and knowledge contained in domains, topology, and networks. A DEKB includes the correspondence, or mapping of above knowledge bases, instead of copies of database schemas, and forms a knowledge pool for sharing data models. The content in a DEKB depends on the purpose of data sharing. For example, data integration on the table level needs ontologies for features and mappings for table schemas; data integration on the system level also involves ontologies for DBMS systems and mappings for system level information. There currently are some mechanisms to obtain the DEKB. Our goal is that with obtained DEKB, a semantic query is translated by knowledge in the DEKB, and the translated query directly search the geospatial database with known destination tables, querying attributes, and filtering values.

TRANSIT SYSTEM DESIGN

A transit data query system was developed using the RDF. The system takes a spatial query in a RDF query language called SPARQL and translates it to spatial queries for Web Feature Services (WFS) and non-spatial RDF queries. The latter queries are handled by a RDF to database mapping program (called D2R Server) that translates the non-spatial RDF query to Structured Query Language (SQL) queries for databases. For

example, the system allows access to geographic data using a query such as "find Route with name 'NorthView' ."

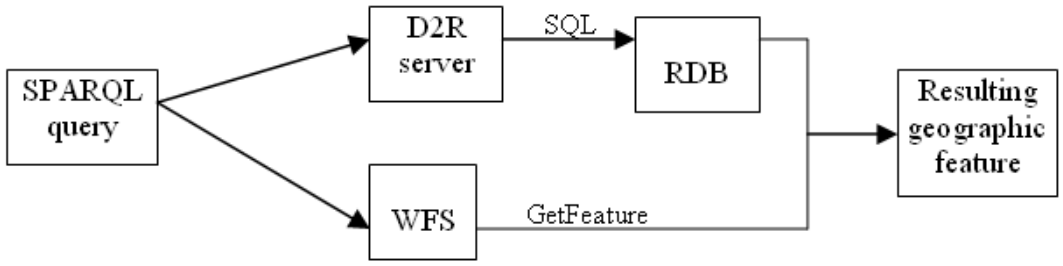

Fig. 4. Query translation strategies used in the RDF-based transit system design.

CONCLUSIONS

In this paper, a conceptual architecture is presented to bridge ontological and geospatial knowledge bases. The goal is to take advantage of semantic power of ontologies and storage and analysis capability of geospatial management systems, avoid data duplication, and construct an integrated system with the advantages of both systems.

REFERENCE

Hammitt, L.C. and Beckert, J., 2007, Ontology-driven information system, U.S. patent 7200563, filed date August 18, 2000, issued date April 3, 2007.

Dobear, C. and Hart, G., 2006, Combining spatial and semantic queries into spatial databases, International Semantic Web Confernece, Nov. 2006, Athens, Georgia, USA.