

Map as a Knowledge Base (MapKB), A Free and Open Source Geospatial Semantic Technology System Prototype

Dalia E. Varanka

U.S. Geological Survey, dvaranka@usgs.gov

Keywords: Geographic data, Linked data, Semantic technology, Knowledge graph

Abstract:

Knowledge Graphs are a form of database representation and handling that show the potential to better meet the challenges of data interoperability, semi-automated information reasoning, and information retrieval. Geospatial knowledge graphs (GKG) are specialized forms of semantic technology that provide coherent spatial context to a domain of information. This paper discusses research toward the development of a prototype GKG based on national topographic databases of geospatial feature instances, attributes, relations, and metadata and contributes to the body of knowledge for other GKGs for the creation of spatial knowledge infrastructure based on topographical and geospatial semantics.

Challenges to GKG development are to capture and represent geographic semantics inherent in the source data through data transformation, design an ontology to enable computation for automated reasoning based on First Order Logic, visualize the data using a cartographic user interface that supports semantic search and other graph data model capabilities, and to integrate data from outside sources primarily through SPARQL Protocol and RDF Query Language (SPARQL) queries. Called the MapKB, the prototype system uses software components to build architecture aligned with semantic specification approaches and is composed entirely of free and open source software. These core components consist of data transformation software, an applied ontology in the form of Web Ontology Language (OWL), a triplestore (a data store for graph-based triples), a SPARQL endpoint, a reasoner, and a user interface. Semantic specification is supported by standardized vocabularies, International Resource Identifiers, and other approaches. MapKB data sources consist of various formats and themes: a gazetteer (flat file), structures represented as point data (volunteered geographic information), government unit boundaries (US Census Bureau), hydrographic data (USGS) and linked open data (Linked Open Data Cloud).

The GKG draws on diverse data formats, themes, and software, but the main interoperability function is performed through the ontology. The ontology consists of thematic modules based on geographic information system (GIS) layers. Sets of rules are established for the appropriate transformation of layer tables to create classes, object and data properties, and annotations by reusing attribute headings and domain and range joins (Wagner and Varanka 2020). The ontology logic is validated by the reasoner and by satisfying competency questions. An extension of SPARQL called GeoSPARQL enables topological spatial relations between features having coordinate geometries. An unofficial release of the software is available at: <https://code.usgs.gov/makb/MapKB>.

In addition to the learning curve to understand semantic technology standards for GKGs, other challenges were to resolve the problems often typical of data silos: attributes and values named in different ways; separate data and metadata sources; missing, redundant, or inflexible relational table model structures for integrating data. The resolution of these issues in the ontology schema pointed to semantic problems of the existing databases and created a basis of discussion for their correction.