[DM-07-080] Ontology for Geospatial Semantic Interoperability

Abstract

It is difficult to share and reuse geospatial data and retrieve geospatial information because of geospatial data heterogeneity problems. Lack of semantic interoperability is one of the major problems facing GIS (Geographic Information Science/System) systems and applications today. To solve geospatial data heterogeneity problems and support geospatial information retrieval and semantic interoperability over the Web, the use of an ontology is proposed because it is a formal explicit description of concepts or meanings of words in a well-defined and unambiguous manner. Geospatial ontologies represent geospatial concepts and properties for use over the Web. OWL (Ontology Web Language) is an emerging language for defining and instantiating ontologies. OWL builds on RDF (Resource Description Framework) but adds more vocabulary for describing properties and classes. The downside of representing structured geospatial data in OWL and RDF languages is that it can result in inefficient data access. SPARQL (Simple Protocol and RDF Query Language) is recommended for general RDF query while the GeoSPARQL (Geographic Simple Protocol and RDF Query Language) protocol is proposed as an extension of SPARQL for querying geospatial data. However, the runtime cost of GeoSPARQL queries can be high due to the fine-grained nature of RDF data models. There are several challenges to using ontologies for geospatial semantic interoperability but these can be overcome through collaboration.

Keywords: GeoSPARQL, interoperability, metadata, ontology, OWL, RDF, SPARQL

Author & citation

Zhang, C. (2019). Ontology for Geospatial Semantic Interoperability. The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2019.4.9.

Explanation

1. Definitions
2. Issues for Geospatial Semantic Interoperability
3. Ontologies for Geospatial Semantic Interoperability
4. Ontology Languages
5. Creating Ontologies
6. Ontology Queries
7. Challenges for Geospatial Semantic Interoperability using Ontologies

1. **Definitions**

Ontology: a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

Web Ontology Language (OWL): a family of knowledge representation languages for

creating ontologies.

Resource Description Framework (RDF): a standard model for data interchange on the Web. RDF extends the linking structure of the Web by using URIs for links and relationships.

SPARQL (Simple Protocol and RDF Query Language): a query language for RDF data.

GeoSPARQL (Geographic Simple Protocol and RDF Query Language): a geographic query language for RDF data.

2. Issues for Geospatial Semantic Interoperability

Many organizations have begun to produce geospatial data for their applications. One of the big challenges for sharing and integrating these geospatial data is heterogeneity among different spatial data. Lack of interoperability, especially semantic interoperability, is one of the major problems facing GIS systems and applications today (Bishr 1998). Geospatial semantic interoperability means computers can share and exchange geospatial data with unambiguous and shared meanings; thus they can make logic knowledge discovery and inference from geospatial data (Kuhn 2005; Janowicz et al. 2012). Therefore, geospatial semantic interoperability is not only concerned with geospatial data syntax but also meanings of data (semantics). Meanings of words may change with different contexts. Synonymy problems occur when multiple words refer to the same concept, while ambiguity issues happen when one word is used to refer to more than one concept depending on its context. Geospatial data users may use different words to describe the same thing or they may use the same words to associate different meanings (semantics) with the words. Computers have difficulty differentiating the meanings of these words with different contexts when ambiguity and synonymy issues occur.

In addition, despite the popularity of geographic contexts, existing web search engines are not designed to help users find geospatial information (Li et al. 2008). Existing web search engines lack semantic knowledge of geographical terminology and associated spatial structures. Therefore, when a user types the name of a place using a typical search engine, many resources associated with the place may not be retrieved. For example, resources relating to places that are inside the specified place may not be found using a typical search engine.

Furthermore, traditional geospatial analyses are mainly based on quantitative measurements or data. And they lack consideration of qualitative geospatial information. However, people often use qualitative descriptions to refer to geospatial locations in their daily life, and understand and express geospatial relations using imprecise qualitative geospatial words such as near, around, or far, based on geospatial-semantic associations from textual and other non-metric information. Cardinal directions such as East, North East, West, and South West are frequently used by people to indicate relative directions among geospatial objects. However, this kind of directional reference is imprecise and it is difficult to identify the exact boundary of geospatial objects using these directional references. Information about geospatial proximity relations (such as A is close to B, X is far away from Y) that is collected by different applications is often vague and imprecise. Current geospatial topological models are not suitable for handling imprecision and vagueness in geospatial topological relations (such as connectivity, adjacency, and intersection among

geospatial objects) expressed in natural language (such as next to, cross, come through, intersect with, split, and surrounded). Yet geospatial references (e.g. east side of the city, west of the river) and geospatial relation describers (e.g. near, far, close) are important for expressing geospatial concepts and computing geospatial proximity and associations.

Therefore, it is necessary to develop ontological approaches for handling “vagueness” in geospatial concepts and relations to facilitate geospatial information retrieval and semantic interoperability.

3. Ontologies for Geospatial Semantic Interoperability

To solve geospatial data heterogeneity problems and support geospatial information retrieval and semantic interoperability over the Web, an ontology is proposed as a formal explicit description of concepts or meanings of words in a well-defined and unambiguous manner (Gruber 1993). An ontology provides a fixed set of concepts whose meanings and relations are stable for computers’ interpretation, inference, and knowledge discovery. An ontology is able to improve communication among computers and enable computer programs to automatically generate transformations among different terminology systems, thus can better support information retrieval over the Web and achieve semantic interoperability among computer systems and applications. A geospatial ontology represents geospatial concepts and properties for use over the Web and is key for successful implementation of the Geospatial Semantic Web. The goal of a geospatial ontology is to make knowledge contained in geospatial applications explicit and express a common understanding of the structures of geospatial information suitable among applications.

Geospatial ontologies consist of geospatial features and geospatial relations. Geospatial features include Point, Line, and Area concepts, and relations include equal, touch, disjoint, intersect, cross, within, contain, near, overlay, connected, in front of, and around, etc. A geospatial ontology model is made up of several elements: geospatial classes/concepts, properties and attributes for these concepts, constraints on properties and attributes, relationships between and among these class concepts.

4. Ontology Languages

There are several ontology languages available to encode an ontology. OWL (Ontology Web Language) is an emerging language for defining and instantiating ontologies (McGuinness and Van Harmelen 2004). OWL is developed to support the Semantic Web and is not limited to a specific application and builds on RDF (Resource Description Framework), but adds more vocabulary for describing properties and classes. RDF/XML (Resource Description Framework/eXtensible Markup Language) is the normative OWL exchange syntax. JSON-LD (JavaScript Object Notation for Linked Data) is overtaking RDF/XML as the more widely used OWL exchange syntax because it is better supported as Intent systems grow and develop. Ontologies created using OWL are normally placed on web servers as web documents, thus they can be referenced by other ontologies and downloaded by applications that use ontologies. In GeoSPARQL, every geospatial features is modeled as a subclass of a top level SpatialObjectclass. SpatialObject class is a subclass of the highest level class Thing in OWL.

Spatial relationships are modeled as object properties on GeoFeatures. OWL is capable of representing geospatial classes made up of Boolean combinations of properties and relationships via set operators on properties (union and intersection).

The downside of representing structured geospatial data in OWL and RDF languages is that it can result in inefficient data access.

5. Creation of an Ontology

It's impossible to create a single ontology containing representations of every term used for different applications. However, it is possible to find some finite set of "primitive" concept representations that can be integrated or combined to create any of the more specific concepts or meanings for different applications. Recently people have begun to actively build ontologies for different applications, for example, an Ontology for Transportation Systems (OTN) has been developed based on GDF (Geographic Data Files) (Lorenz et al. 2005). The USGS (United States Geological Survey) has been working on developing an ontology for the National Map (Usery and Varanka 2012). Many general ontologies have been created in domains such as biology. However, there are few geospatial ontologies for geospatial semantic interoperability. Not many of current efforts on ontology development and management consider the important role of space for our modern computer information systems, and most current ontologies do not consider spatial characteristics of information.

Building ontologies is not an easy task. Ontologies are typically created by a small groups of people such as researchers using ontology tools and editors such as Protégé. These ontology tools and editors are able to help support ontology consistency, visualization of ontologies, and importing existing ontologies. These tools make creating an ontology a little easier. However, it is still an error-prone task to create ontologies using these tools and editors. Even using these tools and editors, it is unrealistic to expect a non-domain expert to create a high-quality ontology without great effort and trial and error. Acquiring or creating an ontology has become bottleneck for achieving geospatial semantic interoperability. It is still a very difficult task to create high quality ontologies using user-friendly and well-designed software programs for real world applications. Converting from the Unified Modeling Languages (UML) to OWL may sever as a way to develop ontologies. But this approach also has many limitations (Zhang et al. 2008).

6. Ontology Queries

Because OWL is based on Description Logics (DL), DL-based reasoners and inference rules are used to collect knowledge bases for automatic spatial queries on the Web. The main benefit from DL is that it can solve the subsumption (subconcept/superconcept) and satisfiability (consistency) problems that often exist in the presenting data. A DL reasoner can check whether two concepts equal, satisfy (consist) or subsume each other (Horrocks and Patel-Schneider 1998). The DL reasoning engines such as RACER, Pellet, or FACT are capable of checking basic consistency of ontology in OWL such as checking class consistency and inferring subsumption hierarchies in OWL. Jena Library can be used to handle OWL format context data and ontologies (Carroll et al. 2004). However, OWL query

engines are still under development. Most of current query engines work only on the RDF level and only simple triple information can be retrieved. These existing reasoners cannot work well for spatial data and relationships. More research is needed to extend these existing reasoners to deal with geospatial data and relationships.

SPARQL was recommended by the W3C (World Wide Web Consortium) for general RDF query over the Semantic Web (Quilitz and Leser 2008). Recently the GeoSPARQL protocol was proposed by OGC (Open Geospatial Consortium) as an extension of SPARQL for querying geographic RDF data (Perry and Herring 2012). OGC GeoSPARQL defines a vocabulary for representing geospatial data in RDF. It defines some query transformation rules that expand a feature only query into a geometry-based query. GeoSPARQL can accommodate both information systems based on qualitative spatial reasoning and quantitative spatial computations. However, due to the fine-grained nature of RDF data model, runtime cost of GeoSPARQL queries can be high because they are dominated by spatial join operations. In addition, because RDF representation of spatial data consists of loosely connected data objects related by object properties, it is very inefficient to process spatial joins due to lack of spatial indices. Because GeoSPARQL queries are much more flexible than database queries and it is difficult to predict which spatial objects should be indexed, pre-computing spatial indices does not guarantee performance improvement. It may be more practical to create spatial indices on demand by implementing extensions to GeoSPARQL.

7. Challenges for Geospatial Semantic Interoperability using Ontologies

Ontology quality and ontology matching are two of the significant challenges undermining geospatial semantic interoperability (Zhang et al. 2015), with ontology quality being the most significant one. As mentioned earlier, creating ontologies manually - even with tool support - may be a difficult and error-prone task. But many issues have yet to be resolved before automatic transformation between the various existing modeling technologies can take place, such as UML to OWL, due to their differences. With more ontologies constantly growing in size, it is increasingly difficult to understand, maintain, and edit ontologies. The online tools for collaborative ontology development may be a good way for creating, maintaining, and editing high quality ontologies. However, many do not have robust web-based interfaces to support collaborative ontology development.

Several issues need to be solved to enable a collaborative ontology development. One essential issue is how to manage conflict resolution for different members of the geospatial community during the collective development process, as differing opinions inevitably cause conflicts. It is, however, not easy to resolve the conflict to guarantee the correct decisions. Further studies are needed as to how to solve this conflict problem and develop a good rating system for ontology editing operations. A search engine may help avoid duplicated concepts, however such search engines are difficult to implement.

Any application that needs multiple ontologies must establish semantic mappings among them to ensure interoperability. Ontology matching is important to solve the geospatial semantic heterogeneity problem by finding correspondences among semantically related ontologies. However, ontology matching is another important challenge for the goal of geospatial semantic interoperability using ontology. Despite its pervasiveness, ontology

matching is still largely conducted by hand in a labor-intensive and error-prone process (Euzenat and Shvaiko 2007; Delgado et. al 2013). Manual ontology matching has become a significant bottleneck for developing large-scale geospatial information management systems. It is important to develop automatic tools to help ontology matching process for the success of different geospatial applications (Cruz and Sunna 2008; Du et al. 2013). Machine learning techniques may be used to semi-automatically create semantic mappings among ontologies.

The lack of sufficient metadata annotations of ontologies is another challenge. Ontologies are always developed by individuals or groups in isolation using different techniques and methods, and a lack of metadata undermines ontology sharing and matching. However, there are no enforced standard conventions for describing contents and context of ontologies. Semantic metadata tools for annotating ontologies and resources should be developed to facilitate ontology discovery and matching over the Geospatial Semantic Web. The World Wide Web Consortium (W3C) has recognized the need to model geospatial metadata in the form of an ontology and noted the problem that metadata have not yet been either modeled or codified. The OpenGeospatial Consortium too has developed standards related to metadata, e.g. OGC I15 (ISO19115 Metadata) Extension Package of CS-W (Catalogue Services for the Web) ebRIM (ebXML registry information model) Profile 1.0 (http://www.opengeospatial.org/docs/is), but these are not consistently implemented.

Finally, nowadays web sites are no longer static web pages serving contents and images. They are now more dynamic, adaptive, and responsive, and this contributes additional uncertainty in ontology creation, maintenance, and matching processes. The use of newer data formats (many of which are schemaless) makes it more difficult to use existing ontology creation, matching, and alignment techniques for the current web applications. Under these conditions, existing approaches for ontology creation, maintenance, and matching need to be modified and new perspectives for solving this problem need to be developed.

In general, it is too cumbersome for only a small group of people to resolve the aforementioned challenges, and achieving the goal of geospatial semantic interoperability via ontologies will require the collaborative work of many people. Crowd sourcing and social approaches will aid in ontology creation and maintenance for sharing and reusing geospatial information and data at the semantic level.

References

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science, 12(4), 299-314.

Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., & Wilkinson, K. (2004). Jena: implementing the semantic web recommendations. In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters (pp. 74-83). ACM.

Cruz, I. F., & Sunna, W. (2008). Structural Alignment Methods with Applications to Geospatial Ontologies Transactions in GIS, 12(6), 683-711.

Delgado, F., Martínez-González, M. M., & Finat, J. (2013). An evaluation of ontology matching techniques on geospatial ontologies. International Journal of Geographical Information Science, 27(12), 2279-2301.

Euzenat, J., & Shvaiko, P. (2007). Ontology Matching, 1st Edition. Heidelberg: Springer.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199-220.

Horrocks, I., & Patel-Schneider, P. F. (1998). DL systems comparison. In Proceedings of the 1998 Description Logic Workshop (DL'98) (Vol. 11, pp. 55-57).

Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial semantics and linked spatiotemporal data–Past, present, and future. Semantic Web, 3(4), 321-332.

Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? In: Spaccapietra, S., Zimányi, E. (eds) Journal on Data Semantics III. Lecture Notes in Computer Science, vol 3534. Springer, Berlin, Heidelberg.

Li, W., Yang, C., & Raskin, R. (2008). A Semantic Enhanced Search for Spatial Web Portals. In AAAI Spring 2008 Symposium: Semantic Scientific Knowledge Integration (pp. 47-50).

Lorenz, B., Ohlbach, H. J., & Yang, L. (2005). Ontology of Transportation Networks. Report from Reasoning on the Web with Rules and Semantics (REWERSE). IST-2004-506779.

McGuinness, D. L., & Van Harmelen, F. (2004). OWL Web Ontology Language Overview. W3C recommendation, 10(10), 2004.

Open Geospatial Consortium (OGC) (n.d.) GeoSPARQL – A Geographic Query Language for RDF Data. OGC Implementation Standard, 40.

Quilitz, B., & Leser, U. (2008). Querying Distributed RDF Data Sources with SPARQL. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds) The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science, vol 5021. Springer, Berlin, Heidelberg.

Usery, E. L., & Varanka, D. (2012). Design and development of linked data from The National Map. Semantic Web, 3(4), 371-384.

Zhang, C., Peng, Z. R., Zhao, T., & Li, W. (2008). Transformation of Transportation Data Models from Unified Modeling Language to Web Ontology Language. Transportation Research Record, 2064(1), 81-89.

Zhang, C., Zhao, T., & Li, W. (2015). Geospatial Semantic Web. Springer.