Encyclopedia of Information Science and Technology, Third Edition

Mehdi Khosrow-Pour
Information Resources Management Association, USA

A volume in the
Geospatial Semantic Web for Spatial Data Sharing

Chuanrong Zhang
Department of Geography, University of Connecticut, USA

Weidong Li
Department of Geography, University of Connecticut, USA

INTRODUCTION

Semantic Web was recently proposed to overcome the semantic heterogeneity problem and provide computers meaningful web content (Berners-Lee et al. 2001). Geospatial Semantic Web is an extension of the current Web, where geospatial information is given well-defined meaning by the ontology so that geospatial contents can be discovered, queried, and consumed automatically by computers. Geospatial Semantic Web aims to add computer-processable meaning (semantics) to the geospatial information on the World Wide Web. Because there are different encodings of geospatial semantics in GIS (Geographic Information System) databases, it is challenging to process requests for geospatial information over the Web. The Geospatial Semantic Web concept was proposed to address the vexing semantic challenges and achieve automation in geospatial web service discovery and execution (Duke et al. 2005).

While Open Geospatial Consortium (OGC) web services provide syntactic ways to encode geospatial information over the Web, they are unable to capture semantics information of geospatial data. Thus, a Web user has difficulty to find an appropriate geospatial data set for a specific task using one of the current search engines, because geospatial data sets encoded using OGC web services lack semantic information and computer programs are unable to understand the meanings expressed by geospatial data contents and requests. However, Geospatial Semantic Web is capable to capture, analyze, and share geospatial information beyond the purely lexical and syntactic level.

This article introduces the spatial data sharing problem and how to make intelligent search and integration of heterogeneous geospatial information by using the Geospatial Semantic Web technologies. We introduce the state-of-art Geospatial Semantic Web technologies such as ontology, ontology-based geospatial web services, ontology-based geospatial web service search engine, and the natural language interface for enabling users to instantly access disparate heterogeneous legacy geospatial data.

BACKGROUND

As GIS has been widely used by a variety of applications, many geographical databases have been developed by different programs and software. However, it is still a big problem to share these geospatial data and use them for the development of GIS applications. Not that spatial data are not available, there is a huge amount of geographical data stored in different places and in different formats; but data reuse for new applications and data sharing are daunting tasks because of the heterogeneity of existing systems in terms of data modeling concepts, data encoding techniques and storage structures, etc. (Devogele et al. 1998).

With the development of open standards, web services emerged for data interoperability over the Web. Within the broader context of web services, OGC web service specifications deal with geographic information on the Internet. OGC web services are evolutionary web standards that enable integration of different online GIS data and location information. With OGC’s web service specifications and technologies, users can “wrap” the existing heterogeneous spatial data into a web service and enable many potential clients to use the service (OGC Interoperability Program White Paper 2001). OGC web services can be treated as a “black box” to perform a task by dynamically connecting...
interoperable service chains for different applications (OGC White Paper 2001).


Unlike the current proprietary commercial GIS formats, the OGC web services support mapping from a wide variety of sources and enable sharing of geospatial data for online information exchanges. The OGC web services provide public open standards for coding and sharing geospatial data. Thus the databases based on the OGC web services can be easily shared and reused (OGC 2003). In addition, the OGC web services also provide a good solution for reducing the costly conversion processes among different format geospatial databases.

ISSUES

Although OGC web services have undoubtedly improved the sharing and synchronization of geospatial information across diverse resources, there are limitations with the currently implemented OGC web services:

First, although OGC web services facilitate data interoperability at the syntactical level via standard interfaces, they cannot resolve data interoperability problems at the semantic level. However, one of the major problems in spatial data sharing and data interoperability is the semantic heterogeneity of spatial data (Bishr 1998; Fonseca et al. 2002; Levinsohn 2000). The OGC web service descriptions only allow the specification of the syntax of basic service contents such as operation metadata, Feature Type list, and filter capabilities, and they provide no semantic descriptions of the meaning of these contents. Two identical XML descriptions may mean very different things depending on the contexts of their uses. In addition, the Web Feature Service specification of the outputs of each call to the service similarly lacks semantic definitions. All defined search operations return results using the same data structure, regardless of what information is requested. For example, a “Building” feature contains a field “Commercial Building,” which is used to describe buildings in commercial areas, and a field “Residential Building,” which is used to describe houses in residential areas. Even if the types of buildings specified in a “Building” file were clearly identified in a “Type” field by the interface designer, OGC WFS descriptions provide no uniform way of enabling such interpretations. It is up to the web service client to recognize the values in these fields, which indicate whether it is a commercial or a residential building.

Second, OGC web services only make it possible to search and access geospatial data by keywords in metadata, but cannot allow content-based searching at the semantic level. Because OGC web service descriptions do not support the semantic specification of service contents and operations, they only can allow requesting semi-structured keyword searches based on the metadata. In addition, metadata also have semantic heterogeneity problems. Different metadata creators may use different names for the same feature. For example, by typing keywords “school” and “Storrs, CT” in a data system implemented using OGC web services in Connecticut, users may get query results of a bunch of feature-level school data such as “Mansfield Middle School” and “E.O Smith High School” data services for Storrs, Connecticut, if their metadata contain exactly these keywords. However, if they use different names for the same feature, it is unlikely that a software program could read and utilize these data services without human assistance, because the OGC web service descriptions provide no means of including representations of the semantics of the provided services. Therefore, with OGC web services it is difficult to perform intelligent content-based search and users cannot correctly utilize the discovered web services without additional human assistance or programming. Further, metadata contain only limited information to allow users to search. Despite the efforts that the geospatial community has put on providing better tools to manage geospatial metadata, content-based searching at the semantic level remains a challenging problem.

Third, without a formal semantic description of OGC web services, it is difficult to allow users and applications to discover, deploy, compose, and synthesize the OGC web services automatically. The lack of an explicit semantic in the XML-based standard OGC web service descriptions proves to be a major limitation in automatic capability matching. It is unrealistic to expect that advertisements and requests of
the OGC web services are equivalent, even if there exists a service that can fulfill exactly the needs of the requester. In order to make OGC web services more practically searchable and ubiquitously available, we need a semantic-based approach such that applications can reason about a service’s capability to a level of detail that permits their automatic discovery and composition; thus, these services can be directly accessed by software applications rather than through users. The OGC web services’ lack of semantic descriptions of the meanings of operations makes it impossible to develop web service clients that can, without human assistance, dynamically find and successfully invoke OGC web services and integrate semantically heterogeneous geospatial data together. The OGC web service descriptions must be interpreted by programmers, who interpret the names of keywords or data contents using other supporting documentations to integrate specific services with their client applications.

Geospatial Semantic Web is able to overcome the aforementioned problems of the OGC web services. The Geospatial Semantic Web technologies, such as Ontology, Description Logic (DL) reasoned, and inference rules, allow OGC web services share and integrate geospatial data contents at the semantic level. Therefore, the systems built on these Geospatial Semantic Web technologies can automatically search and access geospatial data by their contents at the semantic level rather than just by keywords in metadata. The following section introduces the Geospatial Semantic Web solution for users to search, access, retrieve, integrate, and visualize geospatial information through the Internet.

THE GEOSPATIAL SEMANTIC WEB SOLUTION

A framework of geospatial data sharing with Geospatial Semantic Web Technologies is shown in Figure 1. The heterogeneous geospatial databases, which have been created by different agents such as municipal and county governments, police departments, health departments, volunteer organizations, federal agencies, and transportation departments for their routine work using different software systems, can be dynamically discovered, accessed, integrated, and visualized by multiple users. The ontology-based OGC web services, such as Web Feature Services (WFS), Web Map Services (WMS), and Web Coverage Services (WCS), are used to access and manipulate the heterogeneous geospatial databases. The ontologies are used to resolve the semantic heterogeneity of the geospatial data. The final users such as disaster responders, decision makers and environmental planners can query and integrate the heterogeneous geospatial data dynamically by using a friendly graphic interface based on natural language. Geospatial semantic web technologies, such as ontology, ontology-based OGC web services, ontology based search engine, and natural language interface, are the core components in the framework presented in Figure 1. The following subsections introduce these technologies in more detail.

Ontology

The term “Ontology” originated in philosophy and has been used in many different ways. In computer science and information science, an “Ontology” formally represents knowledge using a set of concepts within a domain (Gruber 1993). These concepts can be denoted by a shared vocabulary for their types, properties and interrelationships (Smith and Mark 1998). In Semantic Web, ontologies are used for organizing information and representing knowledge about the world or some parts of it. There are three types of ontologies: Domain ontology, Upper ontology, and Hybrid ontology. A domain ontology models a specific domain and represents concepts in a specific field. For example, geospatial ontologies belong to domain ontology in the geospatial field. An upper ontology is a model of the common objects that can be used across a wide range of domain ontologies. It uses a core glossary, which comprises the terms and associated object descriptions as used in various relevant domain sets, to represent knowledge. For example, “Things” is an upper ontology and can be used by different fields. A hybrid ontology is a combination of an upper and a domain ontology. For example, “Geospatial Things” may be considered as a hybrid ontology because it is combination of an upper ontology “Things” and a domain ontology “Geospatial Features.”

Geospatial ontologies represent geospatial concepts and properties for use over the Internet. In general, Geospatial ontologies include geospatial features and geospatial relations. Geospatial features include Point, Line, and Area concepts. Geospatial relations
define the locations of some geospatial features in relation to other geospatial features. There are three types of spatial relations: topological, directional, and distance relations. Examples of topological relations are “Equals,” “Disjoint,” “Intersects,” “Touches,” “Covers,” “Contains,” “CoveredBy,” and “Within.” Examples of directional relations are “left,” “right,” “behind,” and “in front of.” Examples of distance relations are “nearby,” “far away,” and “close.”

Efforts have been made recently to develop geospatial ontologies for use on the World Wide Web. For example, the W3C Incubator group on geospatial ontologies has worked on developing geospatial foundation ontologies (http://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023). Geospatial ontologies adopt the GeoRSS feature model, which allows the description of rectangles, points, lines, and polygons as geometric representation properties of discerned geographic features. The GeoRSS feature model is consistent with ISO standards. However, it provides a subtle difference in emphasis Web-like feature view or aspect. Geo OWL (http://www.w3.org/2005/Incubator/geo/XGR-geo/W3C_XGR_Geo_files/geo_2007.owl) provides an ontology which closely matches the GeoRSS feature model and utilizes the existing GeoRSS vocabulary for geographic properties and classes.

**Ontology-Based OGC Web Services**

Although OGC web services such as WFS, WMS and WCS improve the sharing and synchronization of geospatial information across diverse resources by resolving the challenges of systemic and syntactic heterogeneity of geospatial data, they cannot resolve semantic heterogeneity problems in geospatial data.
However, since geospatial data are developed by multiple agencies in various fields, differences in the meanings of data in terms of language, code, and message of any other form of representation are unavoidable. Semantic heterogeneity of geospatial data remains one of the biggest challenges for spatial data sharing through OGC web services.

To solve the semantic heterogeneity problem of geospatial data, ontology-based OGC web services are employed in the Geospatial Semantic Web framework. In this approach, the OGC web services are connected to OWL (Web Ontology Language) local ontologies, which can ensure the semantic interoperability of geospatial data. Local ontology refers to semantics used by the different data providers. Because the local ontologies are provided by multiple different data providers, it is necessary to mapping equivalent or related classes and properties in the local ontologies. The ontology server is used to realize this function and it keeps the taxonomy of geospatial terminologies and maintains consistency for different local ontologies. The descriptions of OGC web services are mapped to OWL ontologies to provide a semantically based view of the services, which span from abstract descriptions of the capabilities of the services to the actual data contents that exchange with other services.

To realize automatic search and sharing of geospatial feature data at the semantic level, one important challenge is how to match geospatial features to the predefined geospatial ontology. There are two matching methods: 1) manual matching; 2) automatic matching. Manual matching is time consuming. However, it is not easy to perform automatic matching. Techniques such as the supervised machine learning method may be used to realize complete matching automation after a few initialization steps. Further research is needed to find a good method for automatic matching.

**Ontology-Based Geospatial Web Service Search Engine**

It is important to find suitable geospatial information in the open and distributed environment of current geospatial web services. Geospatial web service discovery is a process of locating, or discovering, one or more documents that describe a particular geospatial web service. The ontology-based catalog service is able to capture the semantics of the user’s query, the web services, and the contextual information that is considered relevant in the matching and discovering process.

Because OWL is based on DL (Description Logic), DL-based reasoning and inference rules are used to collect a knowledge base for the automatic geospatial data-matching engine. One of the characteristics of DL is that it enables systems built on it to infer implicitly represented knowledge from the knowledge explicitly contained in the knowledge base.

An extended DL formalism is employed for spatial relation reasoning. Inference rules are applied to the facts in the knowledge base to determine the spatial relations. The matching engine in the framework may include geospatial data discovery algorithms and composition algorithms. The geospatial data discovery algorithms allow matching a request described with concepts in ontologies with atomic OGC web services provided. Note that the provided OGC web services are also described with concepts in ontologies. However, to precisely fulfill the user’s query, atomic OGC web services may not be enough, and two or more OGC web services may be needed to synthesize the required service. The composition algorithms are developed for this purpose. The composition algorithms allow one to create a workflow of OGC web services by splitting and joining the available web service choices. To increase the efficiency of complex geospatial data discovery, both of the discovery and composition algorithms adopt an index strategy that allows rapidly finding the provided data that match the request.

In general, the reasoning capability and computer-interpretable semantic mark-up in the ontology-based search engine do not restrict geospatial data matching to simple string comparison, but permit more complex semantic matching based on subsumption relationships that can be performed. This search engine makes automatic geospatial data discovery, composition, and synthesis possible at the semantic level.

**Natural Language Interface**

Users in different fields are generally not experts in geospatial data and tools. They may not have much training on how to use geospatial data sets that they wish to query. They usually also have no knowledge about how to perform the semantic query from the ontology-based OGC web services using the ontology-
based search engine. Therefore, a natural language processing (NLP) interface is used to allow users express their needs without having knowledge in ontology or web services. The interface takes input queries using natural language expressions and sends queries to multiple data sources through the ontology-based knowledge base.

There are many challenges in the development of the NLP interface because of the ambiguity of the natural language. It is difficult to transform natural language queries into formal ontology queries because of the difficulty correctly mapping the vocabulary of the natural language to the vocabulary of the knowledge base. Therefore, in the framework the application and domain specific ontologies are used to produce a lexicon for translating user inputs. The lexicon is constructed automatically from the ontologies in the ontology server that creates the verbs, nouns and prepositional phrases with the relations in the ontological structure.

The Stanford Parser (http://nlp.stanford.edu/software/lexparser.shtml) can be used to provide a syntax tree for the natural language query. Based on the syntax tree the sequence of the main word categories Noun (N), Verb (V), Preposition (P), Wh-Word (Q), and Conjunction (C) can be extracted. From the extracted word categories query skeletons can be generated. The generated query skeletons can be used to match with the synonym enhanced triples in the ontologies. The matching is controlled by domain and range information of the ontology.

The lexicon used in the framework is composed of three sources: (1) ontology entities in the ontology server, including ontology classes (concepts), ontology properties (relations), and ontology instances (individuals), which are used to limit the ambiguities and errors in the natural language interactions; (2) general dictionaries, such as WordNet, which are used to enlarge the vocabulary of the ontology and help mapping user vocabulary to ontology vocabulary; (3) application specific synonyms, such as user-defined synonymy words, which are used to define application jargons and abbreviations. The lexicon can be updated in an automatic way from the ontology server for the specific application and domain of use. It can insert the words into the lexicon with the semantic values (classes) that they inherit from the ontologies used by the applications.

By using the NLP interface, users can intuitively find out the requested geospatial features without special GIS, geospatial technology, or programming training. The interface takes advantages of the semantic markup and ontologies to interpret the natural language queries and exposes semantic OGC web services directly to the users.

**FUTURE RESEARCH DIRECTIONS**

The framework presented above introduces concepts and technologies of Geospatial Semantic Web to allow users to directly discover desirable geospatial data from multiple semantic heterogeneous sources. However, since the Geospatial Semantic Web is still at the initial stage of research, there are many challenges in implementing a workable system to query, retrieve, integrate, and visualize heterogeneous geospatial data at the semantic level. The following paragraphs provide some examples of future research directions.

First, further studies are needed to investigate how to create high quality ontologies for different applications. Although some studies have created some specific application ontologies, currently these ontologies are typically built by a small number of people, in most cases by researchers, using ontology tools and editors such as Protégé. Although these ontology tools and editors supporting ontological modeling have been improved over the last few years and many functions are available now, such as ontology consistency checking, import of existing ontologies, and visualization of ontologies, ontology building manually has proven to be a very difficult and error-prone task and becomes the bottleneck of knowledge acquiring processes. For instance, it is unrealistic for non-domain-experts to use these tools to build high quality ontologies. Although transformation algorithms has been proposed by Zhang et al. (2008) to automatically transform the existing UML to OWL so as to avoid errors and provide a cost efficient method for the development of high quality ontologies, there are many issues yet to be resolved due to the differences between UML and OWL.

Second, how to develop a service plan to invoke the execution among services in correct order needs further research. Many applications such as emergency response and disaster management have complex tasks. Usually an atomic web service may not be sufficient to precisely fulfill a user query. Thus two or more web services may be needed to synthesize the required
service. However, the dynamic nature of the availability of web services, the large number of alternative combinations of service choices, and real-time requirements on service composition make dynamic service composition a formidable task. Thus, how to create a work flow of services by splitting and joining the available web service choices is still a research topic.

Third, how to efficiently handle large geospatial knowledge bases in the ontology-based search engine and how to efficiently handle geospatial data reasoning are also waiting for additional study. A search may become slow for data with a large number of geospatial features. While DL is well suitable for the representation of structured or semi-structured attribute information, it becomes complicated if geospatial data is considered. For example, for knowledge bases it would be necessary to be able to compute spatial relationships from the geometry of objects. However, the DL-based system is unlikely to efficiently handle large geospatial knowledge bases. A huge ABox containing all the topological relationships of the spatial data will be quite prohibitive to query without the help of appropriate index structures or optimization techniques.

Fourth, there are many challenges in implementing the NLP interface for different applications. Although a series of discoveries and developments over the past decades have resulted in major progress in the discipline of Natural Language Processing, today many questions still are waiting to be answered to implement an operational system for the real world applications. Here we list a few of the examples: what is the fastest way to find the top-10 semantically related words for each word in the vocabulary of different disasters to fully implement the NLP interface? How to develop algorithms through training on the existing data to utilize multiple levels of annotation, e.g., employ propositional constraints at the same time as employing syntactic constraints? How to accomplish human-like language processing that would require annotation of the pragmatic connotations of language in use for different application purposes?

In addition, several of other concerns such as heterogeneous ontology integration, fault/error handling, reliability, and security issues are still research issues.

**CONCLUSION**

With Geospatial Semantic Web technologies, the heterogeneous geospatial data from multiple sources can be searched, accessed, retrieved, integrated, and visualized dynamically through Internet by different geospatial applications. With the ontology technology, a computer can understand and retrieve the semantically heterogeneous geospatial data, and thus can immediately return useful information without further processing and weeding out irrelevant information. This overcomes the low recall/precision problem faced by the text/keyword based search, which does not capture the underlying semantics of data and forces users to express their query in the vocabulary and syntax of data sources. By providing a natural language interface for accessing geospatial feature information, the solution can dramatically lower the barrier for different applications use the Geospatial Semantic Web. It is unnecessary for users to understand formal ontologies and precisely defined vocabularies.

Although the framework has the obvious advantages, there are many challenges for the real world implementation, and it needs further studies to fully implement a workable system.

**REFERENCE**


## ADDITIONAL READING


**KEY TERMS AND DEFINITIONS**

**Catalogue Service (CS):** The Catalogue Service provides catalogues for OGC web services and supports the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects.

**Geospatial Semantic Web:** Geospatial Semantic Web is an extension of the current Web where geospatial information is given well-defined meaning by the ontology so that geospatial content can be discovered, queried, and consumed automatically by computers.

**OGC Web Services:** OGC web services are evolutionary web standards that enable integration of different online GIS data and location information. The major OGC web service specifications include Web Feature Service (WFS) specification, Web Map Service (WMS) Specification, Web Coverage Service (WCS) Specifications, Web Processing Service (WPS) Specification, and Catalogue Service (CS) Specifications.

**Ontology:** The term “Ontology” originated in philosophy and has been used in many different ways. In computer and information sciences, an “Ontology” formally represents knowledge using a set of concepts within a domain

**Web Coverage Service (WCS):** The Web Coverage Service provides access to potentially detailed and rich sets of geospatial information in forms that are useful for client-side rendering, multi-valued coverage, and input into scientific models and other clients.

**Web Feature Service (WFS):** The Web Feature Service allows a client to retrieve, query, and manipulate feature-level geospatial data encoded in GML (Geography Markup Language) from multiple sources.

**Web Map Service (WMS):** The Web Map Service is capable of creating and displaying maps that come simultaneously from multiple heterogeneous sources in a standard image format.

**Web Processing Service (WPS):** The Web Processing Service defines rules for standardizing inputs and outputs (requests and responses) of geospatial processing services.