Expert Systems with Applications 39 (2012) 12522–12535

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Enhance discovery and retrieval of geospatial data using SOA and Semantic Web technologies

Yong Tian *, Mutao Huang

School of Hydropower & Information Engineering, Huazhong University of Science & Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Service discovery
Web information systems
Information retrieval
Ontologies
Geospatial data
Service-oriented architectures

ABSTRACT

The ability to find and access geospatial data over the Internet through a single tool will be a great benefit for researchers who try to collect relevant data for a particular study. This requires a unified system to bring together diverse data sources so that they can be searched simultaneously in a single environment. Dealing with diversity of data sources and coping with heterogeneities that exist between them are the major challenges in this endeavor. This paper presents the efforts of combining Open Geospatial Consortium (OGC) specifications, Universal Description, Discovery and Integration (UDDI) standards and ontologies to address these challenges. The primary goal is to enhance discovery of web services compliant with OGC specifications and promote access to geospatial information via the OGC services. To this end, this paper introduces an approach for incorporating ontology into UDDI registry and proposes an advertisement algorithm for representing service metadata contained in the OGC capabilities document via UDDI data structures, providing therefore a way to record semantic, descriptive and technical information of OGC services inside UDDI registry. Then the paper illustrates how this information could be used to enhance service discovery with the aid of semantic reasoning. Based on the proposed approaches and service-oriented architectures (SOA) strategies, a web-based prototype system has been developed in order to help users query, access and visualize geospatial data of different types in a unified interface. Several running examples are given to demonstrate the feasibility and effectiveness of the prototype.

1. Introduction

Over the past decades, a huge amount of geospatial data has been collected by different organizations and archived at globally distributed locations, thanks in part to advances in information, satellite and remote sensing technologies. Meanwhile, more data continues to pour in. Scientists and engineers are able to access this valuable data via the Internet sometimes for free. While the Internet opened the floodgates of information, the diversity of the data and data sources also poses great challenges for the users who try to assemble multidisciplinary data sets for a particular study. To overcome diverse nature of the geospatial data and to represent it in a uniform way, syntactic standardization (interoperability) has long been proposed and a number of metadata standards have been developed. As concerns syntactic standardization, Open Geospatial Consortium (OGC), a consortium that works towards interoperability for geographic information, defines a suit of standards for data structures and web services (Lieberman, 2003). These standards provide means to query, access, exchange, deliver and process geospatial data over the Internet in a standard way, no matter what underline platforms are used. Among the OGC standards, Web Map Service (WMS) (Beaujardiere, 2006), Web Feature Service (WFS) (Vretanos, 2010) and Web Coverage Service (WCS) (Baumann, 2009) are most popular and being widely used throughout the world.

Nevertheless, searching for suitable geospatial web services is still a challenging task for end users. Usually, common Internet search engine like Google or Geographic Web portals like the Geospatial One-Stop and the Geography Network (Tait, 2005) are used to find geospatial data or services. However, the users may encounter several problems when using these tools. First, the users may not know what keywords to use or how to fill in interactive forms, especially for novice users (Hochmair, 2005). Second, the search results may be too narrow or too large due to lack of understanding of the semantics of the users' entered search terms as well as lack of spatial inference mechanisms. Third, search results are often presented in the form of hyperlinks, directing the users to providers' web site where the users view or download data. The users may spend much time on eliminating those unsuitable services because they cannot view graphically the actual data content of the services. Finally, after locating suitable data sources, considerable efforts may be consumed on learning relevant specifications in order to extract data overlaid the geographic region of interest.

* Corresponding author. Address: 1037 Luoyu Road, Wuhan 430074, China. Tel.: +86 13476186007; fax: +86 27 87544235.

E-mail addresses: ytian.world@gmail.com (Y. Tian), hmt1973@sina.com (M. Huang).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.061

To overcome these disadvantages, better methods and tools are needed to provide.

Another important issue blocking discovery of geospatial data is a problem widely known as semantic heterogeneity, which occurs when there is a disagreement about the meaning, interpretation or intended use of the same or related terms (Sheth & Larson, 1990). Despite the web service specifications defined by OGC allow for interoperability at systematic and schematic (or structural) level, they largely focus on the syntactic aspects to ensure that it is possible for distributed systems to access, they lack a well defined methodology to describe the thematic dimension of the web services, and tell little about what the severed data represents (Maué, 2009). Most of OGC services provide machine-readable capabilities documents to describe metadata about themselves, but terms or taxonomies used by different organizations to describe the metadata may differ from each other very often. As a result, semantic heterogeneity resulting from the terminological differences becomes a really difficult barrier for information retrieval (Lacasta, Nogueras-Iso, Béjar, Muro-Medrano, & Zarazaga-Soria, 2007). In order for meaningful utilization of OGC web services, a further step has to be taken in providing the user with explicit descriptions, which tells him/her what the geospatial resources (i.e. feature types, map layers, coverage, etc.) actually mean. This calls for a Semantic Reference System (Kuhn, 2003) allowing for semantic annotation of the resources.

The necessities and benefits of semantic annotation have been demonstrated in lots of researches. Annotation of geospatial information refers to the task of attaching meaningful descriptions to the web service and to the served geospatial data or processes through the use of ontology (Maué, 2009). One of the main advantages of the annotation is opening opportunities for more sophisticated query such as matchmaking based on logic reasoning. For example, a casual user looking for data concerning land surface is obviously taking the term ''land surface''. The specialist advertising the data may use more specific term like ''soils'' or ''landscape''. Searching based on keywords matching would yield no results here. In contrast, searching based on semantic reasoning would determine that the advertisement satisfies the user's query because it can recognize that the ''Soils'' is more specific than ''Land Surface''. This approach relies on the philosophy of ontology where the concept ''Soils'' is simply modeled as immediate sub-concept of ''Land Surface''. In computer science, the ontology is a formal data structure that describes a conceptual domain, usually consisting of a set of statements (axioms) that define concepts and relationships between the concepts (Wand, Storey, & Weber, 1999). Ontologies are generally written in formalized languages like Web Ontology Language (OWL) (McGuinness & Harmelen, 2004) to ensure machine-understandable. According to Villa, Athanasiadis, and Rizzoli (2009), ontologies can be used at different degrees of internal complexity and expressive power. The simplest use is as controlled variables. A further step is used as a taxonomy, where concepts are arranged in a generalization/specialization hierarchy. Both the controlled vocabularies and the taxonomy could be used as a basis for semantic annotation. In fact, this usage is becoming common in geosciences and environmental applications (Batzias & Siontorou, 2006; Beran & Piasecki, 2009; Buccella, Cechich, & Fillottrani, 2009; Dupmeier & Geiger, 2006; Horsburgh et al., 2009; Lutz & Klien, 2006).

In addition to semantic annotation, the importance of semantic registry able to store and process both service capability descriptions and semantically annotated information has been quickly identified. This kind of registry is proposed as the main tool for semantic service discovery. It is often built upon extant mature specifications like Universal Description, Discovery and Integration (UDDI). UDDI (Clement, Hately, Riegen, & Rogers, 2004) is currently one of the core standards in the Web service technology stack and a central element feature for service-oriented architecture (SOA) vendor's technology strategy. However, the adoption of UDDI specification in the geospatial communities is slow partially due to its several critical limitations (Gone & Schade, 2007), e.g., its syntax-based search mechanism which may yield coarse results with high precision and recall problems, and the lack of spatial query capability. Since ontology may provide best practices for improving the search capability and alleviating the semantic heterogeneity (Beran & Piasecki, 2009), employing ontology to represent service characteristics and to improve matchmaking ability of UDDI has been the focus of numerous researches in recent years (Kourtesis & Paraskakis, 2008a). Extensions of UDDI standards are also proposed by many researches in order to give UDDI registry the ability of processing requests that contain spatial constraints like the restriction that resulting web services should only offer data covering a specific region.

The objective of this paper is twofold. First, it proposes a solution for enhancing discovery of geospatial services conforming to OGC specifications. The solution relies on a combination of several standards from the domain of Geospatial Web Service and Semantic Web technologies: OGC specifications, for retrieving metadata descriptions of the web services; UDDI, for storing and processing capability descriptions (i.e., descriptive, technical and spatial information) and semantically annotated information about the services; OWL, for modeling a taxonomy pertained to the geospatial domain and for performing fine-grained service matchmaking via semantic reasoning. Implementing a prototype based on the proposed solution as well as SOA principles is the second objective of this paper. The implemented prototype is a web-based system that enables querying OGC services and assembling geospatial data from multiple different sources and of different types and formats in a single interface. Powerful and popular Rich Internet Applications (RIAs) technologies are also leveraged to develop graphic user interface (GUI) of the prototype.

The rest of this paper is organized as follows. Section 2 discusses related work involved in semantic annotation and semantic registries. Section 3 describes the methodology for enhancing semantic discovery of OGC services. The details of the prototype implementation and several investigation results are given in Section 4. Section 5 concludes with an outlook to future work.

2. Related work

In this section, we conduct a general survey of the recent work related to semantic annotations. We also investigate the work that provides solutions to import semantic annotation and other service descriptions into standard-based registry and provide the required discovery mechanisms. Throughout the investigation, particular emphasis is placed on the work targeted to OGC standards.

2.1. Semantic annotation of geospatial information

A number of efforts have been devoted to enrich geospatial web services with semantics by means of ontologies (Cabral, Domingue, Motta, Payne, & Hakimpour, 2004; Klien, Lutz, & Kuhn, 2006; Lutz, Sprado, Klien, Schubert, & Christ, 2009; Martin et al., 2004). The problem of generating semantic annotation of geospatial information is tackled in (Klien, Fitzner, & Maué, 2007). In the work, services descriptions and data schemas are automatically transformed into a formal representing language. Semantic annotations are also generated by making explicit the relationship between the data schema and domain ontology. This process requires definition of mappings from the schema's elements to the concepts in the ontology modeled by WSMO (Web Service Modeling Ontology) (de Bruijn, Fensel, Keller, & Lara, 2005). Specifically, a strategy for