Global polar geospatial information service retrieval based on search engine and ontology reasoning

Nengcheng Chen,[1,2] Dongcheng E,[3] Liping Di,[2] Jianya Gong[1] and Zeqiang Chen[1]

[1]State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University,129 Luoyu Road, Wuhan, China, 430079 (cnc_dhy@hotmail.com;   jgong@lmars.whu.edu.cn;   czq0119@163.com)
[2]Center of Spatial Information Science and System, George Mason University，6301 IVY LANE SU 620 Greenbelt, MD, USA, 20770 (dli@gmu.edu)
[3]Chinese Antarctica Center of Surveying & Mapping (CACSM), Wuhan University,129 Luoyu Road, Wuhan, China, 430079 (edc@whu.edu.cn)

Abstract In order to improve the access precision of polar geospatial information service on web, a new methodology for retrieving global spatial information services based on geospatial service search and ontology reasoning is proposed, the geospatial service search is implemented to find the coarse service from web, the ontology reasoning is designed to find the refined service from the coarse service. The proposed framework includes standardized distributed geospatial web services, a geospatial service search engine, an extended UDDI registry, and a multi-protocol geospatial information service client. Some key technologies addressed include service discovery based on search engine and service ontology modeling and reasoning in the Antarctic geospatial context. Finally, an Antarctica multi protocol OWS portal prototype based on the proposed methodology is introduced.

Introduction

The Open Geospatial Consortium (OGC) has been developing OGC Web Services (OWS)–a seamless framework for a variety of online geo-information processing and location service based on web services technology. OWS references a set of OGC abstract and implementation specifications [Jérôme Sonnet, 2004; Arliss Whiteside, 2005], such as Web Map Service (OGC,2005), Web Feature Service (OGC,2005) and Web Coverage Service (OGC,2006). These specifications support web access to geospatial data and maps through a standard query mechanism.

Antarctica plays a key role in many scientific questions, many of which are related to global climate change. In most of these research activities, the spatial component is crucial. Location is fundamental to field data collection and management and key to supporting advanced data mining in and across extensive spatially enabled data bases. The Antarctic Spatial Data Infrastructure (AntSDI), sponsored by SCAR's Standing Committee on Antarctic Geographic Information (SC-AGI) is responsible for Antarctica spatial data maintenance and sharing through application of OGC specifications. However, the construction of the AntSDI faces the challenge of coordinating access to diverse information resources and services created by many organizations and initiatives. As a result, users face difficulty in locating suitable information and services from these sources.

This paper focuses on the problems of geospatial information service retrieval in the Antarctic context. An approach is proposed through application of a service ontology, semantic web, and web service techniques to register and access Antarctic geospatial resources. The architecture and components of prototype are illustrated according to this semantic web vision. The core technology prototyped includes service discovery based on a search engine, OWS ontology design and reasoning, and the deployment of a service registry based on extended UDDI and OGC catalogue service.

Related work

The "Semantic Web" provides a new perspective on the management and discovery of geospatial information. Ontologies are formalized vocabularies that can be used in semantic processing. The term, "ontology," originates from the field of philosophy as a reference to the nature and the organization of reality; it is a "specification of a conceptualization" (T. Gruber, 1993). To provide formal semantic descriptions of NASA data and scientific concepts, several projects are underway to develop a semantic framework. Described in the OWL language, the ontology within the Semantic Web for Earth and Environmental Terminology (SWEET) [R. Raskin,2006] contains several thousand terms spanning a broad extent of Earth system science and related concepts (such as NASA GCMD, ESML, ESMF, grid computing, and OGC). The SWEET provides a high-level semantic description of Earth system science. The ontology of geographic information metadata (ISO 19115 and FGDC) [Islam L et al, 2003] may provide semantic context to the data description through its explicit association with providers, instruments, sensors and disciplines, and the relationships across these concepts.

In (Zhao P et al,2006; Di L et al,2006), Geospatial data mining ontology oriented toward the research themes in NASA Earth-Sun system is put forward. However, the mapping between the geospatial information service metadata ontology is still in investigation. In particular, the definition of ontology for geospatial information services has not been adopted.

Desktop spatial data mining tools are adopted widely (Deren Li et al, 2001). Fonseca et al. (2002) proposed a

type of semantic framework based on geo-spatial information structures. The SPIN project (May et al, 2003) supported by European Commission is a web based spatial data mining system, which combines existing geographic information system with data mining function into a completely coupled, open and extensible system.

A web-based spatial data mining framework is proposed (Ming Fu, 2004) and a Web Based Spatial Data Mining (WBSDM) platform based on XML/J2EE is designed, that includes a Multi-Agent System based on inspection of a relatively fixed database structure. In (Yuqi Bai et al, 2004), a spatial data search engine "CIHU" is implemented, that locates raster pictures based on place name. Google recently introduced a demonstration of location based web search engine in the USA. Like the Vicinity search tools, it allows the user to specify the name of a place of interest using an address or zip code, which is then matched against relevant documents.

Architecture

Four core components are included in the architecture of the proposed global polar geographic information service (GIService) retrieval (Figure 1).

The Distributed GIService provides access to data layer architecture through services, including a great number of OGC compatible geospatial information services existing in AntSDI (Such as KGIS in Germany, ADD in GB, Map/Feature Catalog in Australia, etc.) or elsewhere. It separates the Prototype engine from the data layer to facilitate improvements and changes to the data structure and/or type of data storage without the need of changing the business logic. An OWS ontology tree is stored as files or in a database. The OWS ontology tree is generated from the knowledge of OGC OWS domain and Antarctica geospatial domain. The ontologic constructs are loaded to memory so it can be used for querying, manipulation and reasoning. The Distributed GIService and OWS ontology tree is composed of the data layer.

A GIService search engine is the core component in the architecture, responsible for the link crawl and OWS service detection through processing of OWS service capabilities and content matching. Service metadata includes information of the service provider, service operation, and service content.An Antarctic domain OWS instance is generated from GIService descriptions (metadata) and OWS ontology tree using the OWS service ontology reasoner.

The extended UDDI Registry includes the access URL and service metadata discovered by engine. This is accessible through a UDDI and OGC catalogue service interface.

The Multi-protocol geospatial service client provides a user interface layer to find, query, and invoke the desired geospatial Web Service.

In the prototype system, the information flow of OWS service search is described as: 1) Open source search

engine, "Nutch" was used to traverse links on web pages and generate a URL database. The indexer uses the content to generate an inverted index of all terms and all pages; 2). The document set is divided into a set of index segments, each of which is fed to a single searcher process; The strings "WMS", "WFS" and "WCS" are parsed as keywords to query the indexed web content; 3) An OGC "getCapabilities" request is sent to the identified "WMS", "WFS" and "WCS" service URL links and the service parameters are processed from the response.
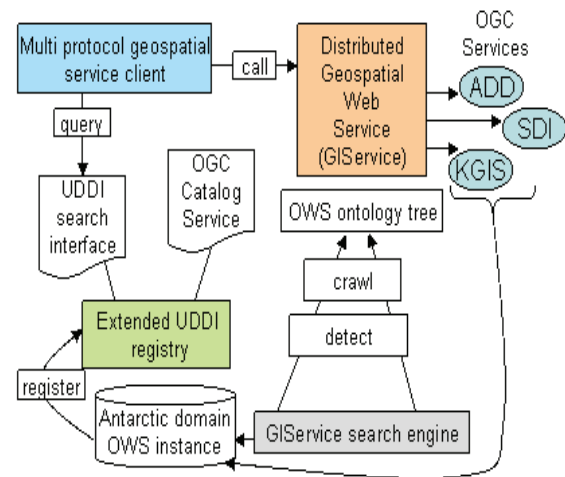

Figure 1. Architecture of prototype system

OWS service ontology modeling and reasoning

In this paper, Protégé 3.1 was used to implement geospatial service ontology. The components of an OGC Geospatial Service ontology tree are listed in figure 2.
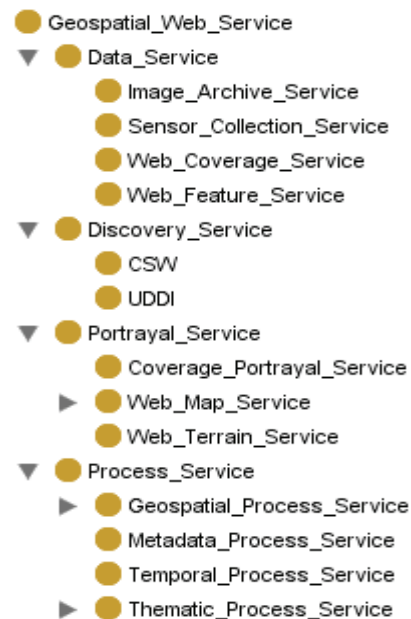

Figure 2. OWS ontology tree

It includes classifications of Portrayal Service, Data Service, Process Service and Discovery Service. Data Service includes Web Feature Service (WFS), Web Coverage Service (WCS), Sensor Collection Service (SOS) and Image Archive Service (IAS). Besides, we can reuse the Antarctica place name ontology developed by Italy. Once populated, the geospatial service ontology enables automatic discovery, invocation and composition of all registered services conforming to the OWL-S specification. The "profile" of OWL-S describes who provides the service, what the service does, as well as other properties of services, and uses the knowledgebase to infer whether or not a particular service is appropriate to a given problem. The "process model" of OWL-S declares the inputs, outputs, preconditions, and effects of a service, and uses the knowledgebase to figure out whether or not a service meets the requirements and conditions to invoke the service.

Since all of the ontology is represented by OWL, the inference engine in the prototype is an OWL reasoner built on Prolog. Ontological information written in OWL or OWL-S is converted into RDF triples and loaded into the Prototype. The engine has built-in axioms for OWL inference rules. These axioms are applied to facts in the prototype to find all relevant entailments such as the inheritance relation between classes that may be not directly revealed in the subclass relationships. An

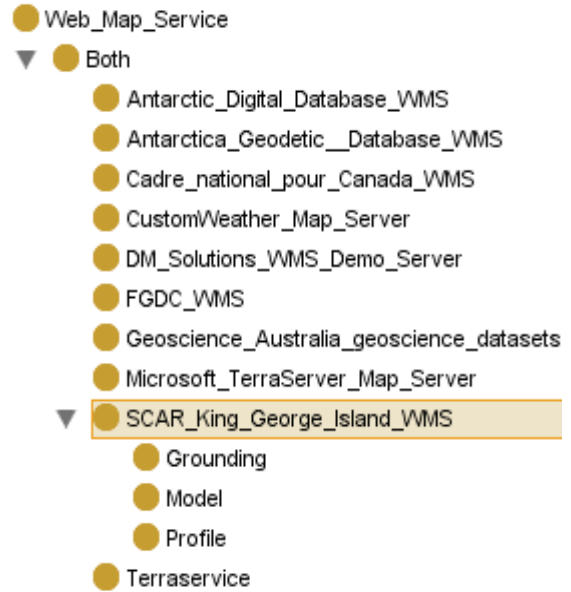Antarctica domain OWS instance is generated and shown in figure 3.

Figure 3. Antarctica domain OWS instance.

Figure 4. Antarctica multi protocol OWS portal (http://polar.chinare.cn/map/)

OWS Service register based on extended UDDI

The UDDI (Universal Description, Discovery and Integration) registry solution was extended to accommodate spatial information in three ways. First, a spatial information services classification was added to reference services as implementing Portrayal, Data, Process and Discovery functionality. Second, information content and format were interpreted to provide additional classification of the spatial data behind the service. Third, the Web Services Description Language (WSDL) file was improved by adding the description of spatial service operation parameters.

The UDDI registry was implemented using jUDDI, an open-source, pure Java web application that can be deployed on any application server or servlet engine that supports version 2.1 or later of the servlet API) and MySQL database. A polar spatial information service hub was constructed in Chinese Polar spatial information portal to access the registry system.

Antarctica multi protocol OWS portal

A multi-protocol OWS service portal (See figure 4) based on the common interface was implemented to support multiple types of geospatial web services (WMS, WFS, WCS) from different implementers supporting the OGC specifications. Through the portal, international polar spatial information services can be published, registered, found, invoked, and integrated. Users can use the portal to find and access distributed polar spatial data through any compatible registered services; these include services from the Chinese polar spatial database using WMS and WFS, the Canadian cyber atlas of Antarctica using WMS, the German King George Island spatial data using WMS and WFS, UK's ADDI data using WFS, the Australian Antarctic Data Centre holdings using WMS, and the USA's USGS polar database using WMS, among others.

Conclusions

This paper demonstrates the value of integrating and extending search engine, registry, and catalogue constructs to include ontology reasoning in order to improve the precision of retrieving polar GIServices. The authors are confident that the mature prototype will contribute a significant capability to the construction of the Antarctic SDI (AntSDI). Future work is anticipated to apply semantics and ontologies in support of more complex user requests.

References

Arliss Whiteside(2005), OpenGIS® Web Services Common Specification. Open GIS Consortium Inc, Canada, 106pp.

Bernard L(2002), Experiences from an implementation Testbed to Set up a National SDI. In proceedings of 5th of the Association of Geographical Information Laboratories in Europe (AGILE), April 25th-27th 2002, Lyon, France.

Deren Li., Shuliang Wang., Wenzhong Shi.,and etc (2001), On Spatial Data Mining and Knowledge Discovery, Geomatics and Information Science of Wuhan University, Vol.26, No.6:491～499.

Di L., Zhao P., Yang W. and Yue P (2006). Ontology-driven Automatic Geospatial-Processing Modeling based on Web-service Chaining, Proceedings of the Sixth Annual NASA Earth Science Technology Conference. June 27-29, 2006. College Partk, MD.

Fonseca F., Egenhofer M., Davis C., and Camara G (2002), Semantic Granularity in Ontology-Driven Geographic Information Systems, Annals of Mathematics and Artificial Intelligence, Vol.36, No.1-2:121～151, 2002.

Islam, L. Bermudez, B. Beran, S. Fellah, M. Piasecki (2003), Ontology for Geographic Information – Metadata. http://loki.cae.drexel.edu/~wbs/ontology/iso-19115.htm.

Jérôme Sonnet (2004), OWS 2 Common Architecture：WSDL SOAP UDDI. Open GIS Consortium Inc, Canada, 76pp.

May M., Savinov A (2003), SPIN! -An enterprise architecture for spatial data mining, LECTURE NOTES IN ARTIFICIAL INTELLIGENCE, 2773: 510-517.

Ming Fu (2004), The Research of Web-Based Spatial Data Mining, Doctor's Degree Dissertation, Central South University.

OGC 02-069 (2005)，OpenGIS implementation Specification：Web Map Service，http://www.opengeospatial.org/docs/01-068r3.pdf.

OGC 02-058 (2006) ，OpenGIS implementation Specification：WFS，https://portal.opengeospatial.org/files/?artifact_id=7176.

OGC 03-006r3 (2006), OpenGIS implementation Specification：WCS, https://portal.opengeospatial.org/files/?artifact_id=3837.

R. Raskin(2006), Enabling Semantic Interoperability for Earth Science Data. http://sweet.jpl.nasa.gov/EnablingFinal.doc.

T. Gruber(1993), A translation approach to portable ontologies, Knowledge Acquisition, Vol. 5, Issue 2, 1993, pp. 199-220.

Yuqi Bai, Chongjun Yang (2004), Research on Spatial Information Search Engine, Journal of China University of Mining and Technology, Vol.33, No.1:90～94.

Zhao P., Di L(2006). Semantic Web Service Based Geospatial Knowledge Discovery. Proceedings of 2006 IEEE International Geoscience and Remote Sensing Symposium. July 31- August 04, 2006. Denver, Colorado, USA.