ONTOLOGY BASED APPROACH FOR GEOSPATIAL SEMANTIC WEB

TRAN Thai Binh¹, Thilo WEHRMANN¹, Steffen GEBHARDT², Verena KLINGER¹, Juliane HUTH², VO Quoc Tuan³, Claudia KUENZER¹

¹German Aerospace Center, German Remote Sensing Data Center, Muenchner Strasse 20, 82234 Wessling
²Department of Remote Sensing, University of Wuerzburg, Germany, Hubland, 97074 Wuerzburg, Germany
³Department of Land Resources, College of the Environment and Natural Resources, Can Tho University

binh.tranthai@dlr.de/ ttbinh@vast-hcm.ac.vn; thilo.wehrmann@dlr.de; steffen.gebhardt@dlr.de; verena.klinger@dlr.de; juliane.huth@dlr.de; vqtuan@ctu.edu.vn; claudia.kuenzer@dlr.de

KEY WORDS: Semantic web, geospatial semantic web, ontology

ABSTRACT: Semantic web facilitates a new generation of web applications. It has been applied in many fields so far, e.g. for medicine, bioinformatics, social web. Up to now, the semantic web is well known as an effective infrastructure to enhance visibility of knowledge on the web. It provides a common interoperable framework in which information is given a well-defined meaning such that data and application can be used by machines (e.g. reasoning) for more effective discovery, automation, integration and reuse across various application, enterprise and community boundaries. The semantic web is based upon an ontology approach. However, in the field of geospatial domain, the ontology approach for web application is not very popular. In this paper, we first briefly present the Resource Description Framework (RDF) and Schema (RDFS) as a metadata model for conceptual description or modelling of information that is implemented in web resources as also Web Ontology Language (OWL) for authoring ontologies which is the primarily supported language for semantic web, and then review the recent ontology approach in the geospatial domain. We will discuss about the lack of efficient ways to find user requested, suitable information in a data warehouse system. Based on the discussion, we will present an ontology based approach for a water-related information system. Our approach aims to provide an intelligent information query tool based on different applications/purposes. The proposed system will provide information appropriate to the user requirements with less or non prior knowledge of the underlying data.

1. INTRODUCTION

Since it was invented in 1990 by Tim Berners-Lee, World Wide Web (WWW) has dramatically grown. With a modest amount from the beginning, WWW consists of billions websites nowadays becoming a very huge and heterogeneous data warehouse. As a result, searching for particular information becomes more difficult. User always faces the problem that searching result is not reasonable or does not relate to the keyword query [12].

A growing number of individuals and groups from academia and industry have been evolving the Web into another level - the Semantic Web [3]. It is “a web of data that can be processed directly or indirectly by machine” [2] and brings a higher degree of automation in exploiting data in a meaningful way. Semantic web facilitates a new generation of web applications. It has been applied in many fields so far, e.g. for medicine, bioinformatics, social web. Up to now, the semantic web is well known as an effective infrastructure to enhance visibility of knowledge on the web. It provides a common interoperable framework in which information is given a well-defined meaning such that data and application can be used by machines (e.g. reasoning) for more effective discovery, automation, integration and reuse across various applications, enterprises and community boundaries. Ontology is applied because it constitutes formal and explicit specification of shared conceptualization of the domain [8].

In parallel with the development of the Semantic Web, the Geospatial Semantic Web is initiated recently. It was identified as an immediately considered research priority for many research groups as UCGIS (University Consortium for Geospatial Information Science) [5] and OGC (Open Geospatial Consortium).
2. CURRENT ISSUES
Web service for geospatial information exists for quite a long time. To make it useful for end-user, the collected data is mostly arranged following the Content Standard for Digital Geospatial Metadata [10] of the Federal Geographic Data Committee (FGDC), or the ISO19115 standard [15]. These standards define the content, quality, condition, origin, and other characteristics of spatial data. In general, the current information systems are organized in spatial-, thematic- and temporal aspects ([13], [14], [17]). With the approach of Gebhardt et al. for the WISDOM system, the user can explore data by using search criteria with thematic, regional and temporal attributes for necessary data. Results can be browsed or downloaded for further analyses. The systems can manage various spatial and non-spatial datasets and their distinct aspects. However, it also has some difficulties, especially for novice users, who may not know which keywords to use, have too little information on how to fill in interactive forms, or find it difficult to estimate how many filter criteria should be utilized [9].

Further, data is managed in cross-related tables causes other problems. The relation of datasets of different hierarchical schemes apparently causes some redundancies. It is hard to manage the datasets if they belong to different thematic groups and to maintain them [6]. In additional, Decision Support System (DSS) may support operational management and strategic policy-making and planning by monitoring and analysing the “as is” situation, as well as the forecasting of future situations and disasters (e.g. early warning system) [16], [7]. So the information system must integrate some auxiliary processes. The problem for this situation is that the system must understand the matching between process and data, only then it can respond to a certain query. These aforementioned issues are caused by lacking of meaning of data and processes.

3. STATE OF TECHNOLOGY
As aforementioned, for replying the appropriate information to user queries, the system must understand the meaning of data and the installed processes. Ontology has a critical role describing the meaning of the data in which the computer can understand data to apply meaningful data processing automatically. It is not only useful for sharing understanding, but also evolving as basic for improving data usage, achieving semantic interoperability, developing advanced methods for representing and using complex metadata, correlating information, knowledge sharing and discovery.

The architecture of semantic web is depicted in fig.1 which describes the main layers of the semantic web design [1]. In this architecture, RDF is a basic data model that identifies objects (“resources”) and their relations to allow information to be exchanged between applications while conserving meaning. RDFS is a semantic extension of RDF, it describes the properties of generalization-hierarchies and class of RDF. OWL is on top of RDF and RDFS to add vocabulary to explicitly represent the meaning of terms and their classes’ relationships. For more details about the architecture of semantic web read “A Semantic Web Primer” [1]

3.1 RDF AND RDFS
RDF is a W3C recommendation framework [19] for description the Internet resources. It can be considered as the metadata of the data on the web. Information is presented by subject-predicate-object triples in RDF. The subject always is a resource; the object could be a resource or a literal. RDF uses the XML syntax, hence exchange between different operating systems is easily possible. RDF describes not only the meaning of data, but also the relationship between them [20].
RDFs is an extension of RDF, it can be used to define application-specific classes and properties for RDF, while RDF describes resources with classes, properties, and values. RDFs does not provide actual application-specific classes and properties, but the framework to describe it. Classes in RDFs are much like classes in object oriented programming languages. This allows resources to be defined as instances of classes, and subclasses of classes. The Simple Protocol and RDF Query Language (SPARQL) is a SQL-like language for querying RDF data. For expressing RDF graphs in the matching part of the query, TURTLE syntax is used [20].

3.2 OWL
OWL was designed to provide a common way to process the content of web information (instead of displaying it). It is also a W3C standard for semantic web since 2004; it is built on top of RDF. OWL was designed to be interpreted by computers and used for processing information on the web. RDF, OWL is also using XML syntax. There are three sublanguages: OWL Lite, -DL, and -Full. Each of the sub-language is an extension of the previous one, any legal statement of OWL Lite is a legal statement of OWL DL, and any legal statement of OWL DL is a legal statement of OWL Full, but that is not true in the other direction. (W3C)

3.3 ONTOLOGY BASED APPROACH
There are three ways to apply ontology based approaches [21].

Single approach has one global ontology; all the information sources are related to only one ontology. It can be considered as a hierarchical, terminological database. It may be consisted of several specialized ontologies. It is useful when all information sources to be integrated provide nearly the same view on a domain and has minimal ontology commitment. Changing the information sources is susceptible, because it needs changes in the global ontology and in the mapping to the other information sources.

Multi approach describes each information source by separate ontologies, so it simplifies integration and supports changes in source. There is no shared vocabulary between ontologies, so inter-ontology mapping is needed to communicate between information sources. This approach is difficult to compare different source ontologies as it does not have a common vocabulary.

Hybrid approach is a combination of the aforementioned approaches. Each source is described by its own ontology and shared vocabulary is built to share the basic terms of a domain. The advantages are new sources can easily be added, it supports acquisition and evolution of ontologies and source ontologies are comparable because of shared vocabulary. But existing ontologies can not easily be reused; designers have to redevelop from scratch.

4. RELATED WORKS
The previous research on ontologies mainly focus on integrating different geospatial web services [11] to provide knowledge-based search engine for searching the appropriate information. The ontology was introduced to solve the heterogeneity of data due to the syntax, the structure and the semantics. It was used to describe the meaning of information itself in a particular domain, not to describe what/why the information is needed for.

To deal with the system which has thematic data belonging to different classifications, M. Lutz (2009) [12] proposed a hybrid ontology approach that has a Semantic Translation Specification Service (STSS). The STSS uses the ontology based reasoner as an interpreter. Reasoner is a piece of software able to infer logical consequences from a set of asserted facts or axioms and the inference rules are commonly specified by means of an ontology language [Wiki]. This approach just focuses on translating the information from different classification/definition systems; it focuses only on one thematic field.
Another approach to discover information in geodata services by N. Athanasis et al. (2008) uses multiple ontologies to describe the domain. The data was described in three schemes, the first one describes the data classification, the second one describes the natural hazard, and the third one is about the mandatory elements of the ISO19115 metadata standard. When users search for appropriate information with three criteria the system uses RDF resource description to find the suitable data and send it back to the users’ browser. This approach just uses ontology to describe the meaning of data and does not have any process to provide value-added data/information products.

Marius Podwyszynski (2009) [13] presents an approach, on that data is described based on the related application. The system is divided into two domains, application domain and data domain; the application domain describes the application/phenomenon user working with, data domain describes the properties of data. The two domains are related with each other through a measurement component. Users can search data by the application they are interest in without any knowledge about low level data. But the system only describes Earth observation data.

The queried information, in most cases, will be used for further application/purpose. These applications normally need more than one dataset. The current approaches of the WISDOM system are sufficient to explore the information but they still need effort to search the information because the users have to search several times for each particular dataset and related document (for example users want to analyse the land cover affected by flood. They have to search for water mask datasets, land cover datasets from satellite images, province/region area, legal documents and planning programs of the current region etc). There is no system available with operational data processing capability using an ontology approach to provide the requested information.

5. ONTOLOGY BASED APPROACH FOR THE WATER-RELATED INFORMATION SYSTEM

The proposed approach applies for the information system within WISDOM project (http://www.wisdom.caf.dlr.de) that consists of huge amount of data from many different research fields, from geographic data as shape files and satellite imageries, sensor data for water quality, water level, and household to statistical, text based data as water knowledge, livelihoods and knowledge management. The collected data are heterogeneous because of different formats, different scales and areas of interest and different disciplines.

The system aims to provide a tool to search for appropriate information related to water. Because different disciplines have different aspects of the same phenomenon in the real world, the best solution is to describe data based on the meaning that people think about phenomena. Through surveying the user requirements, these following typical requests should be taken into account 1) certain thematic data related to certain region/province at certain time, 2) the data changing (statistical) during certain period of time due to a certain thematic.

To answer such kind of request (1), the system just needs the data itself which is described by using ontology, but for the request (2), system must have some processes for analysis and to produce reports. In our approach, the system is divided into four different domains (fig. 2): Application Domain, Data Domain, Processing Domain and Observed Object Domain and auxiliary concepts for measurements or spatial relationships. We apply a hybrid approach in the system, so these domains share the common vocabulary in water-related domain.
**Application Domain:** this domain describes the users, the tasks and their business processes. The task is described with the relationship with business process and the observed object domain via the phenomenon/thematic. The user is described in detail by capacity, activity, or responsibility. This domain relates with observed object domain via the phenomenon/thematic and the processing domain via the process that is defined by business process.

**Data Domain** describes the data collection with dataset sub classes. This domain describes the physical properties of a dataset (e.g. data model, structure, property, file format, service, geometric resolution, spatial representation and spatial relationship), temporal entity and observed objects that the dataset is related to as well. The data domain has a relationship with auxiliary concepts for spatial relationships and measurements. It includes the ontology for time, unit, space extracted from SWEET ontology [19].

**Processing Domain** describes necessary processes in detail to transform data to information. Utility, input and output of process are also described.

**Observed Object Domain** describes physical and non-physical objects that relate to the water domain. The mean of observations is also described in this domain, it is divided into two subclasses: earth observation and ground stations. These two groups describe how real objects are observed / measured in detail.

**Auxiliary concept** has the subclasses space, numerical entity and units. These ontologies are extracted from SWEET ontology to describe the measurement for things. All concepts in the system which measure phenomena have relationships with this auxiliary concept.

Regarding the work flow (Fig.3), the user sends 1) a request to the application domain, it assigns a task. The intrinsic task is to find the matching data and proper business process. The business process will be defined due to the semantic description by application ontology. Then the process chain component will be sent to 2) the processing domain. The correct data used for the process relate to the thematic or phenomenon component that is described in the observed object domain. The data query will be sent to 3) the observed object domain with the semantics of thematic/phenomenon of data. The observed object domain then identifies observed object by using the reasoner based on thematic/phenomenon aspect. It will search and retrieve the appropriate object. The certain information of dataset will be identified by data domain with 4) the transferring of information from observed object domain. The next step will be 5) the transaction of information of dataset from the data domain to the processing domain. Within the processing domain, these processes were defined by business process workflows in 6) like a processing chain. The output of a previous process is the input for next one. Finally, the
information will be sent back to the user in an appropriate output form. It should be visualized via WMS, WFS or a report in PDF document format. For a typical data querying request (1), the system should run only step 1), 2), 3), 4) and send the dataset back to the user. The result will cover one or more datasets due to the requested thematic and application.

6. CONCLUSION
This paper presents a new approach to apply an ontology to a water-related information system. We describe the system into four domains using a hybrid approach. The system also includes some business processes to analyse data to provide appropriate information for users. This study is still experimental but the proposed results should permit an innovative approach for the geospatial semantic web to increase the utility of data. This study focuses only on the water-related domain but the architecture can be used and extended to another domain.

7. ACKNOWLEDGE
The WISDOM project is funded by the German Federal Ministry of Education and Research (BMBF) and the Vietnamese Ministry of Science and Technology (MOST).

References