How to Submit Proof Corrections Using Adobe Reader

Using Adobe Reader is the easiest way to submit your proposed amendments for your IGI Global proof. If you don't have Adobe Reader, you can download it for free at http://get.adobe.com/reader/. The comment functionality makes it simple for you, the contributor, to mark up the PDF. It also makes it simple for the IGI Global staff to understand exactly what you are requesting to ensure the most flawless end result possible.

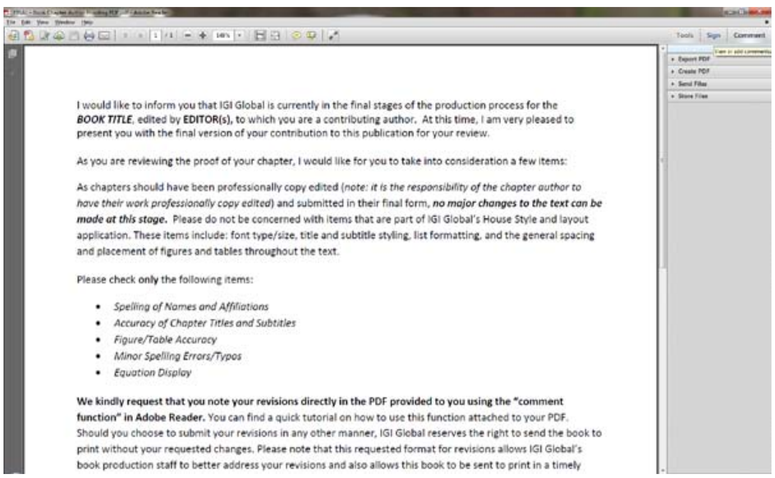
Please note, however, that at this point in the process the only things you should be checking for are:

Spelling of Names and Affiliations, Accuracy of Chapter Titles and Subtitles, Figure/Table Accuracy, Minor Spelling Errors/Typos, Equation Display

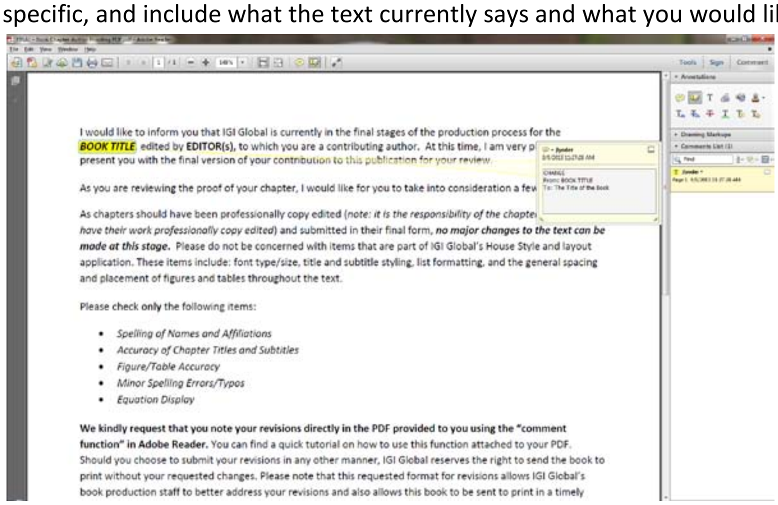
As chapters should have been professionally copy edited and submitted in their final form, please remember that ***no major changes to the text can be made at this stage*.**

Here is a quick step-by-step guide on using the comment functionality in Adobe Reader to submit your changes.

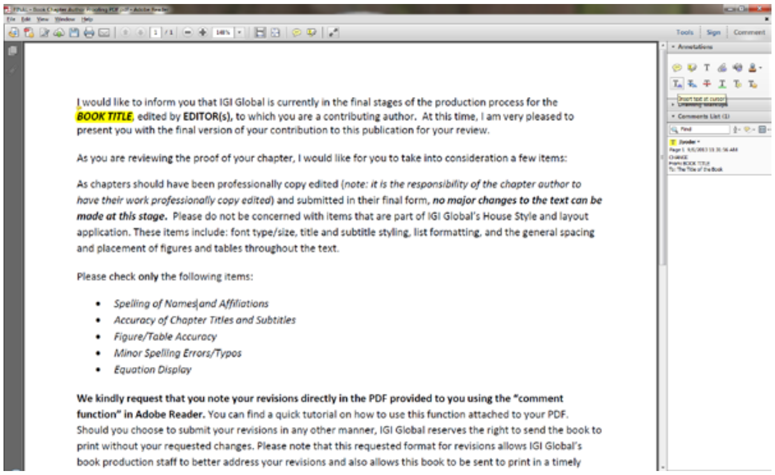
1. Select the **Comment** bar at the top of page to View or Add Comments. This will open the **Annotations** toolbar.

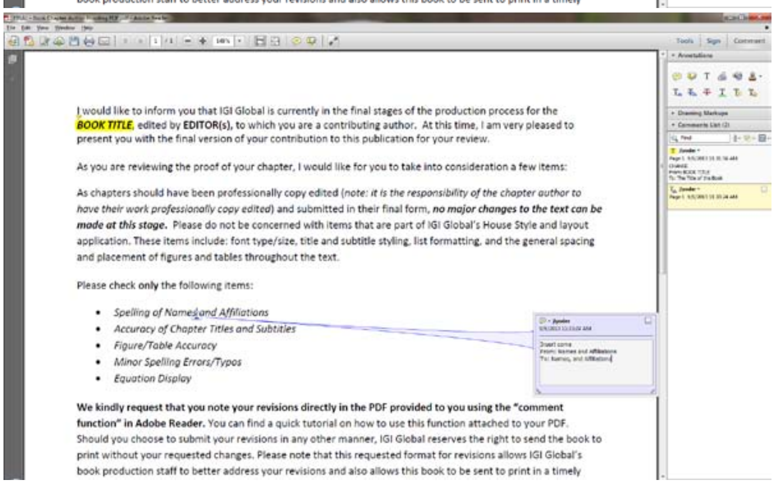


2. To note text that needs to be altered, like a subtitle or your affiliation, you may use the **Highlight Text** tool. Once the text is highlighted, right-click on the highlighted text and add your comment. Please be specific, and include what the text currently says and what you would like it to be changed to.

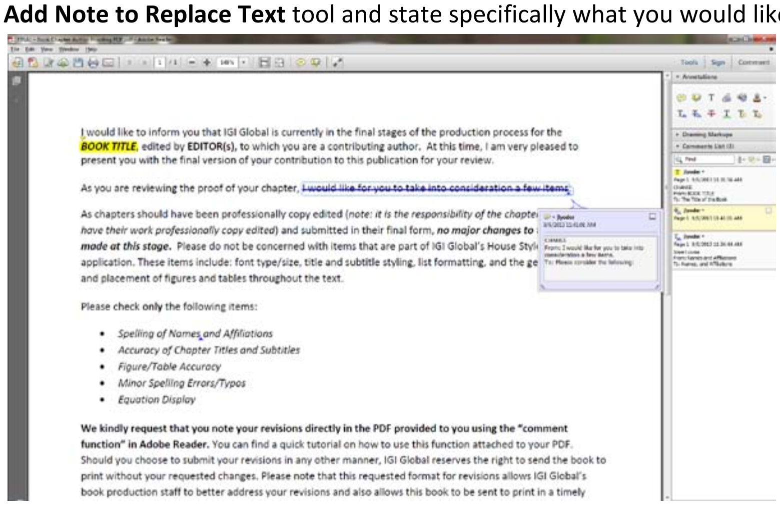


3. If you would like text inserted, like a missing coma or punctuation mark, please use the **Insert Text at Cursor** tool. Please make sure to include exactly what you want inserted in the comment box.





4. If you would like text removed, such as an erroneous duplicate word or punctuation mark, please use the **Add Note to Replace Text** tool and state specifically what you would like removed.



IGI PUBLISHING
WWW.IGI-GLOBAL.COM

International Journal of Applied Geospatial Research

April-June 2014, Vol. 5, No. 2

Table of Contents

Research Articles

1 **Online Flood Information System: REST-Based Web Service**
Xiannian Chen, West Virginia University, Morgantown, WV, USA
Xinyue Ye, Bowling Green State University, Bowling Green, OH, USA
Michael C. Carroll, Bowling Green State University, Bowling Green, OH, USA
Yingru Li, Auburn University, Auburn, AL, USA

11 **Built Environment and Driving Outcomes: The Case for an Integrated GIS/GPS Approach**
Xiaoguang Wang, Central Michigan University, Mount Pleasant, MI, USA
Lidia Kostyniuk, University of Michigan, Ann Arbor, MI, USA
Michelle Barnes, University of Michigan, Ann Arbor, MI, USA

30 **Geographic Information System Effects on Policing Efficacy: An Evaluation of Empirical Assessments**
Yan Zhang, Sam Houston State University, Huntsville, TX, USA
Larry Hoover, Sam Houston State University, Huntsville, TX, USA
Jihong (Solomon) Zhao, Sam Houston State University, Huntsville, TX, USA

44 **Using Semantic Search and Knowledge Reasoning to Improve the Discovery of Earth Science Records: An Example with the ESIP Semantic Testbed**
Kai Liu, George Mason University, Fairfax, VA, USA
Chaowei Yang, George Mason University, Fairfax, VA, USA
Wenwen Li, Arizona State University, Fairfax, VA, USA
Zhipeng Gui, George Mason University, Fairfax, VA, USA
Chen Xu, George Mason University, Fairfax, VA, USA
Jizhe Xia, George Mason University, Fairfax, VA, USA

59 **Spatial Intelligence for Regional Analysis**
Chenfeng Zhang, East China University of Science and Technology, Xuhui, Shanghai, China
Shuming Bao, University of Michigan, Ann Arbor, MI, USA
Bing She, Wuhan University, Wuchang, Wuhan, China
Xinyan Zhu, Wuhan University, Wuchang, Wuhan, China
Xu Zhang, Wuhan University, Wuchang, Wuhan, China

The *International Journal of Applied Geospatial Research* is indexed or listed in the following: ACM Digital Library; Bacon's Media Directory; DBLP; Google Scholar; INSPEC; JournalTOCs; Library & Information Science Abstracts (LISA); MediaFinder; SCOPUS; The Standard Periodical Directory; Ulrich's Periodicals Directory

Using Semantic Search and Knowledge Reasoning to Improve the Discovery of Earth Science Records:
An Example with the ESIP Semantic Testbed

Kai Liu, George Mason University, Fairfax, VA, USA

Chaowei Yang, George Mason University, Fairfax, VA, USA

Wenwen Li, Arizona State University, Fairfax, VA, USA

Zhipeng Gui, George Mason University, Fairfax, VA, USA

Chen Xu, George Mason University, Fairfax, VA, USA

Jizhe Xia, George Mason University, Fairfax, VA, USA

ABSTRACT

Web resources exploration is increasingly driven by semantic web technologies with automated processing. Earth science communities generate large amounts of datasets described in hundreds of millions of metadata records. It is critical to discover the accurate data from the millions of data records based on the end user's searching intent. However, the big challenge is how to ensure that catalogs and Spatial Web Portals can understand end user's intents. To enable portals effectively 'understand' the meaning of user's queries and to provide a better searching experience for end users, we collaborated with Earth Science Information Partners (ESIP) to develop such a capability through a semantic Testbed. We implemented a reasoning engine using similarity calculations to facilitate the meaningful discovery of Earth science data and to improve the accuracy of searching results.

Keywords: *Cyberinfrastructure, Earth Science Information Partners (ESIP), Geospatial Platform, Knowledge Reasoning, Semantic Search,*

DOI: 10.4018/ijagr.2014040104

INTRODUCTION

Earth science communities generate and publish datasets and services described in metadata records. To promote the broad sharing of the geospatial data, services and other resources among public users and government, researchers proposed the Spatial Web Portal (SWP; Yang et al., 2007), which can be considered as an interface to geospatial cyberinfrastructure (Yang et al., 2008), in which the mechanisms for Earth science data storage, indexing, editing, searching, visualization and analysis are provided through an interactive web interface. For example, the FGDC Virtual Arctic Spatial Data Infrastructure (SDI), which is established upon the Service-Oriented Architecture (SOA), has incorporated most available Arctic WMSs for online service chaining and map integration (Li et al., 2010; Li et al., 2011). We built for the intergovernmental GEO ("Group on Earth Observations," 2011) the GEOSS (Global Earth Observation System of Systems) clearinghouse (http://clearinghouse.cisc.gmu.edu/geonetwork) to facilitate the discovery, access, and utilization of Earth observation data, information, tools and services using standardized metadata. By July 2012, 133 remote datasets or services and 167 K metadata have been registered/harvested by the GEOSS Clearinghouse. The ever-increasing resources in national catalogs and clearinghouse pose great challenges for effective resource discovery.

Traditional searching tools, built upon keyword matching technology, are weak in understanding user behavior and providing the most relevant results. Success in searching engines of SWP is not only a matter of quantity of the resources but also the quality of the resources found. Two factors are always used to evaluate the performance of the process of Earth Science records discovery using SWPs: precision and recall. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved ("Precision and Recall," 2011). 1) Users of the Earth science data and information are hindered by syntax mismatches between users and providers (Raskin & Pan, 2005). With millions of geospatial data, services and other resources, there is a big challenge for the catalogs and SWPs to search the most relevant records to help users discover the geospatial information effectively. 2) Normally, SWPs discover Earth science records by matching text using search terms input online by end users. It is difficult for SWPs to understand the meanings of the search terms and do the extensive discovery. Therefore, both the precision and recall are important and should be considered when improving the efficiency of records discovery.

The 21st century witnessed the emergence of the semantic Web (Berners-Lee, 2001) for web resources exploration with a focus on automated processing. The goal of the semantic Web is to augment the current World Wide Web (WWW) with a highly interconnected network of data that can be easily exploited and processed by both machines and human beings. Thus, the semantic Web is designed to make Web data more meaningful so that it can be understood, interpreted, manipulated, and integrated. To this end, W3C proposed a series of formal specifications to specify how Web resources could be modeled, interpreted and presented. Some of these include Resource Description Framework (RDF), RDF Schema (RDFS) and Web Ontology Language (OWL). Some semantic discovery researches based on ontology matching and integration have been introduced to Earth Science (Zhang et al., 2010 a). By formalizing such semantics of user query behavior and modeling them in these standardized machine languages, the semantic web can help machines further improve the performance of a search engine.

This paper reports our research to improve the discovery of Earth science records based on the semantic Web using a case study of ESIP semantic testbed (Yang et al., 2008). The research problem we are trying to address is "Among all the results returned, which ones fit best a user request?" For example, a query of "Natural resource WMS" will return many different records and it becomes extremely difficult for users to pick the best match. Therefore, it will be helpful if the system can evaluate the relevance between the Earth science records and "Natural Resource WMS" to rank the results. This paper presents our research on using semantic similarity calculations for results ranking.

In Section 2, we present a literature review. Section 3 presents the system architecture of the ESIP semantic testbed. Section 4 illustrates the semantic technology used in the ESIP testbed which includes ontology, semantic search, and semantic similarity evaluation. Section 5 introduces the protocol and use case, which shows the feasibility of using semantic search and knowledge reasoning to improve the discovery of Earth science records. The paper ends with conclusions and future research discussions.

LITERATURE REVIEW

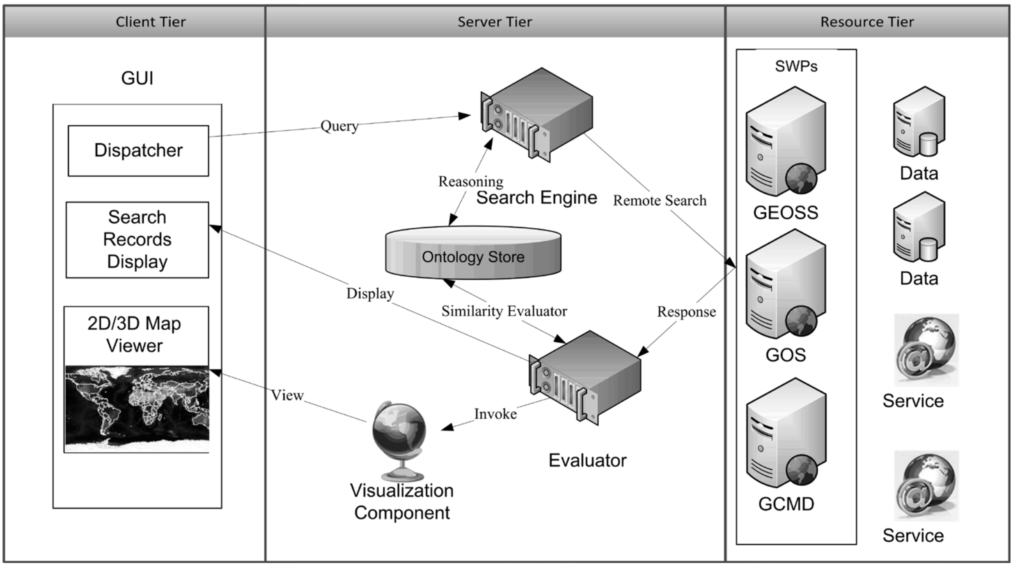
To improve the accuracy and relevance of the results to user intent, a number of research efforts have been undertaken. CSW (Catalogue Services for Web, Nebert, 2007) is commonly used for SWPs to publish and share Earth science records. CSW supports the publishing and search against collections of descriptive information (metadata) for data, services, and other geospatial resources. Catalog services are used to support the discovery of registered information resources within a collaborating community. Some SWPs adapt CSW to publish records, e.g., GEOSS clearinghouse (Liu et al., 2011) and GOS support CSW 2.0.2 standard for metadata search (ESRI, 2007). OpenSearch is another mechanism to search for geospatial data and other resources (OpenSearch, 2011). Some SWPs adapt OpenSearch to publish records, e.g., Global Change Master Directory (GCMD, http://gcmd.nasa.gov/).

Semantic search is a mechanism used by researchers to incorporate the related information from a conceptual or knowledge perspective to retrieve more relevant results or more targeted results with the addition of understanding the user's intent and the contextual meaning of terms as they appear in the search context. Sematic search is used in both the Web and closed systems ("Semantic Search," 2011). The availability of large amounts of structured, machine understandable information of a subject that is captured by the semantic Web offers some opportunities to improve traditional search (Guha et al., 2003). The main purpose of semantic search is using semantics to enable machines to understand the meaning of information on the Web.

In 2005, Di proposed a framework for automatic geospatial knowledge discovery in the Web service environment (Di, 2005). The framework can provide: (1) standards-based automated geospatial data and services discovery and access; (2) domain knowledge driven intelligent geo-object decomposition for geo-tree/workflow construction; (3) automated geospatial Web service chaining, binding, and execution based on the geo-tree/workflow; and (4) management of workflows and geospatial models. Zhang et al. (2007) proposed a framework for the geospatial data sharing based on semantic Web technologies. The framework uses Geospatial semantic Web, OGC Web services and SOA for enabling disparate GIS to share and integrate geospatial information at the semantic level in a cost effective way. This framework allows the sharing of geospatial data from heterogeneous databases at the semantic level over the Web through ontologies and OGC Web services (Zhang et al., 2010b, Zhang et al., 2010c). Zhao et al. (2007) presented a method to enable ontology query on spatial data available from WFS services and on data stored in databases. These works introduces major improvements over other semantic discovery studies in both the consideration of domain context and the automation of processing.

Getting more relevant records is a step to increase the recall of a search. However, increasing the precision is also needed to improve the effectiveness of a search. To increase the precision, a relevance calculation should be conducted. There are several ways to implement the calculation: a) A natural way is node-based approach, which is to evaluate semantic similarity in a taxonomy based on evaluating the distance among the nodes corresponding to the

Figure 1. System architecture



items compared — the shorter the path from one node to another, the more similar they are. Given multiple paths, one takes the length of the shortest one (Resni, 1995). b) Edge-based distance approach is a more natural and direct way to evaluate semantic relevance in a taxonomy by estimating the distance (e.g., edge length) between nodes which correspond to the concepts/classes being compared. Given the multidimensional concept space, the conceptual distance can be conveniently measured by the geometric distance between the nodes representing the concepts (Jiang & Conrath, 1997). Obviously, the shorter the path from one node to the other, the more similar they are (Jiang & Conrath, 1997).

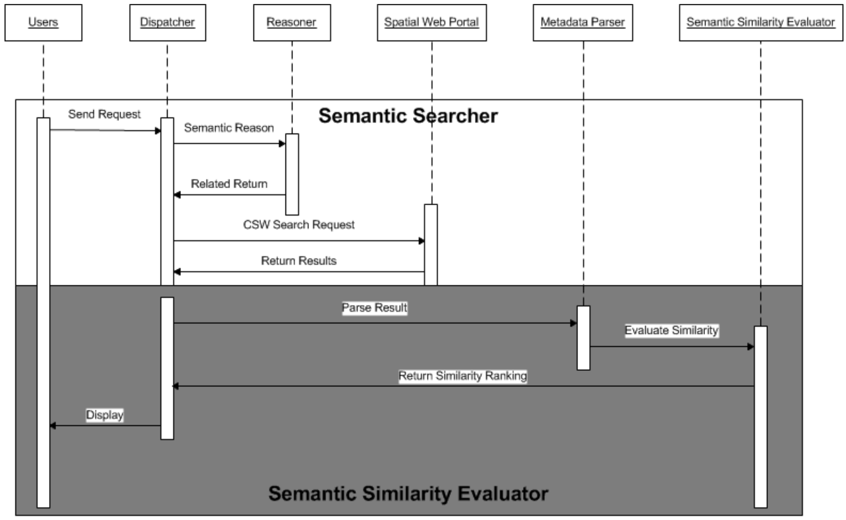
Based on our research of Earth science ontology and the catalog or clearinghouse based data discovery platforms, such as GOS and the GEOSS clearinghouse, we conducted this research to utilize semantics to generate more search results and improve the precision by ranking the search results through the semantic similarity calculations. The system is integrated into the ESIP semantic testbed.

SYSTEM ARCHITECTURE AND WORK FLOW

Figure 1 shows the system architecture of the ESIP semantic testbed, which adopts a three tier models: client tier, application tier and data tier (Ramirez, 2000). Each tier has its own specific functionalities.

At the client tier, a query dispatcher is used to parse users' query terms and send to the semantic search engine. For example, if we search for "Natural Resource WMS", the query will be parsed to "Natural Resource" and "WMS". "Natural Resource" can be queried from ontology store, while "WMS" can't be done in this way. The Graphical User Interface (GUI) is used to display the results for the users and contains two parts: 1) records display part lists the data key fields, such as the title, abstract, key words, and other for metadata. AJAX (Garrett, 2005) is used to exchange data with the server tier, and update parts of the interface page (Li et al., 2010). 2) a 2D/3D viewer provides a visualization interface for some Web services, such as OGC (Open Geospatial Consortium) WMS ("Web Map Service," 2011) and OGC WFS ("Web Feature Service," 2011).

Figure 2. Workflow



The application tier is also known as the logic tier or the middle tier ("Multitier architecture," 2011). It contains three parts: 1) The search engine is the core of the semantic discovery and is used to describe the query phrase before deriving additional information with any optional ontology information based on the axioms and rules. It is also used to send query requests to remote SWPs. 2) The similarity evaluator is used to evaluate the relevance between the returned records and the user's original query. 3) The visualization component provides visualization function to display OGC WMS and WFS. Both the search engine and similarity evaluator are implemented based on the ontology store.

Resource tier includes remote resources, especially, catalogs which publish and store metadata of Earth science data and provide open APIs for remote search (such as, CSW and OpenSearch). Some catalogs, such as GEOSS clearinghouse and GOS, provide CSW to users for the search. In this paper, we use GEOSS clearinghouse as a remote resource to implement the discovery.

Figure 2 illustrates the workflow of the query process, which can be divided into two separate sub-workflows including semantic query and similarity evaluation denoted as white and grey parts respectively in Figure 2.

In the semantic query, when a user sends a search request to the ESIP semantic search portal, a reasoner will conduct semantic reasoning and return relevant concepts. For example, if we send "Natural Resource WMS" to the ESIP testbed, the reasoner will return "Natural Resource, mineral, metal, biomass, fossil fuel, gas hydrate, soil". The dispatcher will compose one CSW GetRecords request with filter tag when it gets the semantic results and send the GetRecords request to the SWP.

In the semantic similarity evaluator part, the dispatcher will send search records to the metadata parser to get the important fields, which include the fields having high contribution for the similarity evaluation. Then, the semantic similarity evaluator module will calculate the distance between each important field and query terms. At last, an algorithm will be used to calculate the overall relevance between metadata and query terms.

SYSTEM ELEMENTS

Ontology

Ontologies are used to represent the knowledge in the semantic Web and formally define a common set of terms that are used to describe and represent a domain knowledge ("OWL," 2004). By defining shared and common domain theories, ontologies help people and machines communicate concisely—supporting semantics exchange, not just syntax (Maedche & Staab, 2001). An ontology is crucial for describing the semantic content of data, to complement the syntactic content that appears in Earth Science Markup Language (ESML) (Ramachandran et al., 2004) descriptor files or other metadata descriptions. Usually, ontologies are the vocabulary and the formal specification of the vocabulary, which can be used for expressing a knowledge base (KB).

The Resource Description Framework (RDF) is the core data representation format for the semantic Web. RDF was originally created in early 1999 by W3C as a standard for encoding metadata and uses URIs and XML schemas to describe things. As an extension of RDF, OWL stands for Web Ontology Language, and is part of the growing stack of W3C recommendations related to the semantic Web ("OWL," 2004). In general, OWL can be defined as a language which extends RDF schemas with other new constructs. Currently OWL is the most popular language to use when creating ontologies (Yu, 2010).

Ontology is an important part for the semantic Web. The semantic Web's success and proliferation depends on quickly and cheaply constructing domain-specific ontologies (Maedche & Staab, 2001). Currently there are many ontology editing tools that can help us construct OWL ontologies such as: Protégé, a java based and open source ontology editor ("Protégé(Software)," 2011), CmapTools Ontology Editor (COE), a java based ontology editor based on CmapTools (ihmc, 2011), and OWLGrEd, a UML style graphical editor for OWL ("OWLGrEd," 2011).

Using the Protégé ontology editor, we developed an ontology to describe natural resources in the Earth science based on the SWEET ontology, where all the terminologies are defined in different facets, including phenomena, property, substance, and Earth realm (Raskin & Pan, 2005). An excerpt from the natural resource ontology is shown below. It shows that the "Mineral" is subclassof "Natural Resource", and soil is subclassof "Natural Resource" and "Mixed Substance".

Semantic Reasoning

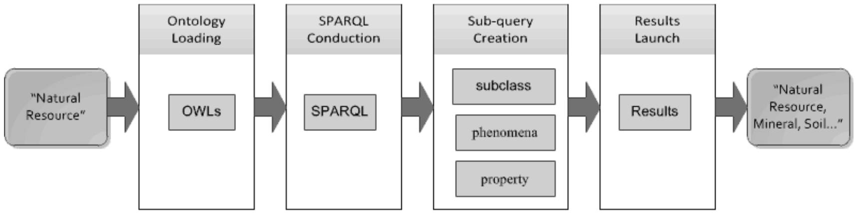
Reasoning is the process to find new knowledge or concept based on prior knowledge. Semantic reasoning is the core component of the semantic search engine [Li et al., 2008] and implements logical consequences based on a set of inference rules. Mostly the set of rules is described using ontologies. With the semantic reasoning and query terms, relevant conceptions and conclusions can be identified from the ontology store.

Description Logics (DLs) are a family of logic-based knowledge representation formalisms that are tailored towards representing the terminological knowledge of an application domain in a structured and formally well-understood way (Baader et al., 2005). DLs allow users to define the important notions (classes, relations, objects) of the domain using concepts, roles, and individuals; to state constraints on the way these notions can be interpreted; and to deduce consequences such as subclass and instance relationships from the definitions and constraints (Tessaris et al., 2009). Because OWL is based on DL, we can use a DL-based reasoner to process a set of semantic queries.

We use the Jena Semantic Web Framework ("Jena," 2011) for semantic reasoning. Figure 3 shows the semantic reasoning procedures of "Natural Resource":

1. Ontology loading.
2. SPARQL (SPARQL, 2008) query conduction, SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed

Figure 3. Semantic resoning procedures of "Natural Resource"



as RDF via middleware. In this step, DL-based query should be converted SPARQL queries.

3. Sub-queries creation with Jena: the following sub-queries can be constructed: phenomena query, subclass query, and property query. Using these sub-queries to traverse the ontologies, we can get a graph of all concepts related to the query terms. Doran, Palmisano and Tamma (2008) present SOMET graph traversal algorithm for ontology module extraction. We built on this algorithm to develop the ontology traversal algorithm described below in Box 1.
4. Results launching: the query results are combined in an appropriate manner to get expanded and more specific information.

Search Conduction

We use the OpenGIS Catalog Services Specification standards including the CSW organization and implementation for the discovery and retrieval of metadata for geospatial data and geoprocessing services. The CSW contains some required operations such as GetCapabilities, GetRecords, DescribeRecord and GetRecordbyId. GetCapabilities is used to describe the catalog service instance. GetRecords is used to search catalogue content and retrieve all or some members of the result set. GetRecords adapts a filter whose syntax is described in the response of GetCapabilities to restrict the search

Box 1.

```
graphTraversal (O,e, R)
INPUT A set of ontologies O, a concept e, a set of mapping
rules R
Output the graph g of all concepts related to e
Initializ a empty set g
For each mapping rule r ∈R
    S = a set of triples apply (e, r) to O
    If S is not empty
       For each s ∈S
          If s ∉g
              Insert s into g
              graphTraversal (O,s, R)
          End If
       End For
    End If
 End For
```

Box 2.

```
<ogc:Filter>
   <ogc:Or>
   <ogc:PropertyIsLike escapeChar="\" singleChar="?" wild-
Card="*">
    <ogc:PropertyName>AnyText</ogc:PropertyName>
    <ogc:Literal>*Natural Resource WMS*</ogc:Literal>
   </ogc:PropertyIsLike>
   <ogc:PropertyIsLike escapeChar="\" singleChar="?" wild-
Card="*">
    <ogc:PropertyName>AnyText</ogc:PropertyName>
    <ogc:Literal>*Mineral WMS*</ogc:Literal>
   </ogc:PropertyIsLike>
   </ogc:Or>
</ogc:Filter>
```

results, and also the operation adapts a more complex scalar predicate by using the logical operators AND and OR (Nebert, 2007). From the GEOSS clearinghouse, a filter of GetRecords is illustrated as below in Box 2.

Previous codes show a CSW filter. Logical operator OR is used to combine the search conditions. The filter is used to search the records which contain "Natural Resource WMS" or "Mineral WMS" in any text of metadata.

Semantic Similarity Evaluation

There are several alternative ways to define similarity such as node-based approach and edge-based approach (Jiang & Conrath, 1997). The semantic similarity evaluation algorithm we used is based on the semantic distance (Rips & Shoben, 1973). We adapt the edge-based approach by using metadata discussed in the process of semantic similarity evaluation. Then, we make a set of additional assumptions about similarity. Similarity measure can then be derived from these assumptions.

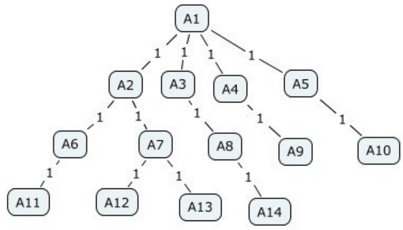
Assumption 1: For geospatial records, we distinguish them through features. Every feature has its contribution for the relevance. GEOSS clearinghouse is chosen as the search source where most records are stored under ISO-19139 (FGDC, 2011) format. There are three basic features chosen in this paper: title, descriptiveKeywords and abstract. Then we give an algorithm to calculate the relevance between metadata and query terms as follows:

$$Sim(r,q) = \alpha^* Sim(k,q) + \beta^* Sim(t,q) + \gamma^* Sim(a,q) \quad (1)$$

Where α, β, γ are the contribution of each feature. The contribution can also be described as the weight. Sim(k, q) represents the similarity between descriptiveKeywords feature and query phrase; Sim(t, q) represents the similarity between title and query phrase; Sim(a, q) represents the similarity between abstract and query phrase. Sim(r, q) is the overall similarity between the record and query phrase. If the query phrase appears in the descriptiveKeywords, we consider the feature of descriptiveKeywords is the main factor to impact the similarity and give the highest weight to β. If the query phrase doesn't appear in the descriptiveKeywords, we will enlarge the α or γ. However, the weights need to change according to users' feedback.

Assumption 2: Every feature has its similarity to user's query phrase and the similarity can be calculated by measuring distance in the ontology. If we consider the ontology as a network, the simplest form of determining the distance between two elemental concept nodes, A and B, is the shortest path that links A and B,

Figure 4. Ontology network



i.e., the minimum number of edges that separate A and B (Rada et al., 1989).

$$Sim(A,B) = \frac{e}{Dis(A,B) + e} \quad (2)$$

$Dis(A,B)$ is the measure distance. The similarity between A and B will get lower if the distance between A and B gets longer. When $Dis(A,B)$ equals zero, the similarity is 1. In addition, e is the modification value that represents the distance when Similarity is 0.5.

The measure distance can be calculated from the ontology. If B equals A, the measure distance is 0; if B is the child of A, the measure distance is 1.

Figure 4 shows an ontology network in which we consider the edge distance between two linked nodes as 1. Hence, the distance between A11 and A10 is 5 in this network. If we consider e as 2, the similarity between A11 and A10 is 0.28.

PROTOTYPE IMPLEMENTATION

ESIP Testbed

To test the effectiveness of this approach, we built the ESIP testbed prototype to search against GEOSS clearinghouse. We used Protégé to build the Earth ontologies. In the prototype, a simple box is used to input the search conditions. When we search "Natural Resource WMS", the semantic reasoner would parse the query to several related conceptions. After the semantic reasoning, a CSW GetRecords request will be combined and sent to the GEOSS clearinghouse through Post request support by HTTP protocol ("POST(HTTP)," 2011). After the semantic search, the ESIP testbed will get the response from the GEOSS clearinghouse. The response is in the XML format. We used JDOM (Hunter, 2002) to parse the response and then get the useful features such as title, abstract, and descriptivekeywords. The semantic similarity evaluator will be used to calculate the similarity between search results and query terms. In addition, some other factors will affect the relevance between the records and user's request, e.g., the completeness of metadata and reliability.

In this prototype, we added some correction values to rank the results. If a result has low adequacy and reliability, we will decrease the ranking of the result.

Figure 5 illustrates the GUI of the ESIP testbed prototype. The left side includes the search results from GEOSS clearinghouse with ranking based on similarity calculations. The title and abstract information are listed in the GUI to show the records. If the result is OGC WMS or OGC WFS, there are 2D and 3D visualization buttons under the record. The 2D visualization function uses OpenLayers ("OpenLayers," 2011) to display the maps in a Web page. The 3D visualization function uses NASA World Wind (NASA, 2011) to show the maps in visually rich 3D. The right side provides the tree and the graph of semantically related concepts.

Figure 5. ESIP testbed prototype

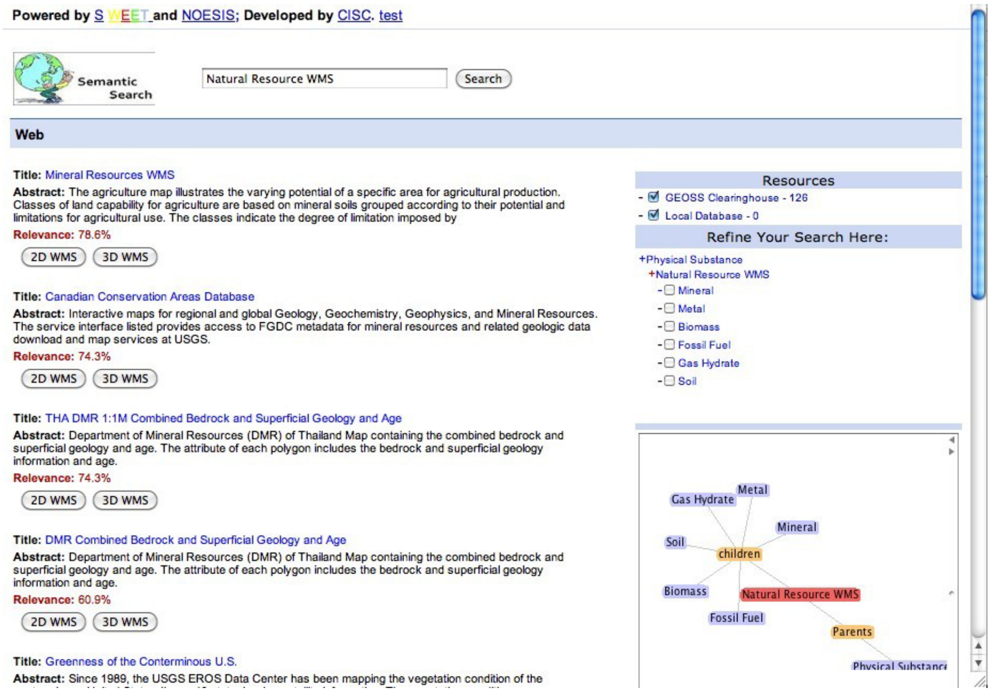


Figure 6. Results wihout semantic search from GEOSS clearinghouse interface

Contrast Before and After This Technology

Figure 6 shows the search results of "Natural Resource WMS" without semantic search from GEOSS clearinghouse local search interface.

To search the records from GEOSS Clearinghouse, we used the advanced search interface in GEOSS clearinghouse (Liu et al., 2011). Compared to the 126 records returned in ESIP Testbed prototype, there are only 5 records returned without using semantics. In the GEOSS clearinghouse, we only get records which contain "Natural Resource" and "WMS" in the text such as the first record "Canadian Conservation Areas Database"; the "Natural Resource" exists in the abstract of this record. The results show that taking the semantics into the discovery of Earth science records can make a difference in improving the recall not only in theory but also in practice. In addition, the results are ranked based on their relevance to the search phrase in the ESIP testbed prototype. The first record is "Mineral Resource WMS"; and its relevance is 78.6%. Although this record does not contain "Natural Resource" in its text, it contains "Mineral Resource" in its title. According to the semantic search and semantic similarity evaluation discussed before, "Mineral Resource" is a related concept of "Natural Resource" and this record has relevance 78.6% to the search phrase. Hence, through the ranking based on the relevance, the semantics are useful to improve the precision of the discovery.

CONCLUSION AND FUTURE WORKS

This paper discusses a semantic Web testbed for facilitating better discovery of Earth science records and demonstrates through integration with the GEOSS clearinghouse. Semantic reasoning methodology is leveraged to support the search engine. In addition, the testbed provides an edge-based similarity approach to evaluate the relevance of Earth science records and query phrases. The research helps us use the geospatial data more effectively by improving the search with easier, faster, and more accurate results. The research results is being integrated into GEOSS, Geospatial Platform, and NASA Spatial Web Portal to enable the better discovery, access, and utilization of geospatial resources to enable spatial cloud computing (Yang et al., 2011) and a geospatial cyberinfrastructure (Yang et al., 2010) for EarthCube (NSF, 2011).

The paper provides a general semantic evaluation method for the discovery. However, every user has their own opinion about the relevance between records. For some users, some records have special importance and the users think these records have higher relevance than other records. Even though both relevance feedback and semantic retrieval have received extensive attention separately, feedback techniques have not yet been developed for semantic retrieval (Yang et al., 2005). A feedback technique is useful in the semantic discovery. In these techniques, users can recommend some records to have high relevance and to give the feedback to the semantic discovery system. Using the feedback solution, the discovery system can interact with users directly at the semantic level.

In addition, the research is being further expanded to cover more domains of Earth sciences, such as the nine areas GEOSS focuses on: natural and human-induced disasters, the environmental sources of health hazards, energy management, climate change and its impacts, fresh water resources, weather forecasting, ecosystem management, sustainable agriculture, and biodiversity conservation (GEO, 2011) for benefiting a broader audience. Another aspect is to consider the data quality and service quality to improve the search results with better quality by expanding the factors to be considered for data quality. Data quality is an important component of the metadata normally associated with authoritative data sets (Goodchild, 2009). For the discovery of Earth science records, quality is a major concern because it determines the limits of use for any data set (Paradis & Beard, 1994).

We also plan to utilize users' search behaviors, including preferences, habits, feedback and others to improve the results. Through the analysis of these parameters, more efficient discovery functions will be developed and these parameters are helpful for ontology engineers to improve the ontology base.

ACKNOWLEDGMENT

Research reported is supported by ESIP Federation, NASA (NNX12AF89G), NSF (IIP-1160979), and Microsoft Research Connection's Earth, Energy, and Environment Program.

REFERENCES

Baader, F., Horrocks, I., & Sattler, U. (2005). Description logics as ontology languages for the semantic web. *Mechanizing Mathematical Reasoning*, 228–248.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*, 34–43. doi:10.1038/scientificamerican0501-34 PMID:11396337

Di, L. (2005). A framework for developing web-service-based intelligent geospatial knowledge systems. *Journal of Geographic Information Sciences*, *11*(1), 24–28.

Doran, P., Palmisano, I., & Tamma, V. (2008). SOMET: Algorithm and tool for SPARQL based ontology module extraction. In *Proceedings of the International Workshop on Ontologies: Reasoning and Modularity (WORM-08)*.

ESRI. (2007). *Using the geospatial one-stop operational portal API – draft*. Retrieved Setermber 13, 2011, from http://www.fgdc.gov/grants/2009CAP/GOS_APIDocumentationFinal1.1.pdf

FGDC. (2011). *Geospatial metadata standards*. Retrieved Setermber 11, 2011, from http://www.iso.org/iso/catalog_detail.htm?csnumber=32557

Garrett. J. (2005). Ajax: A new approach to Web applications. *AdaptivePath.com*. Retrieved October 13, 2011, from http://www.adaptivepath.com/ideas/ajax-new-approach-web-applications

GEO. (n.d.). *GEOSS*. Retrieved September 13, 2011, from http://www.earthobservations.org/geoss.shtml

Goodchild, M. F. (2009). The quality of geospatial context. In K. Rothermel, D. Fritsch, W. Blochinger, & F. Dürr (Ed.), *Quality of context: First International Workshop*, QuaCon 2009, Stuttgart, Germany. Berlin, Germany: Springer.

Group on Earth Observations. (n.d.). In *Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Group_on_Earth_Observations

Guha, R., McCool, R., & Miller, E. (2003). Semantic search. In *Proceedings of the 12th International Conference on World Wide Web (WWW '03)* (pp. 700-709).

Hunter, J. (2002). JDOM and XML Parsingk, *Oracle Magazine*. Retrieved Setermber 23, 2011, from http://www.jdom.org/docs/oracle/jdom-part1.pdf

ihmc. (2011). *Ontology creation for the rest of us*. Retrieved Setermber 23, 2011, from http://www.ihmc.us/groups/coe/

Jena – A semantic Web framework for Java. (2011). Retrieved Setermber 19, 2011, from http://jena.sourceforge.net/

Jiang, J., & Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*.

Li, W., Yang, C., Nebert, D., Raskin, R., Houser, P., Wu, H., & Li, Z. (2011). Semantic-based service chaining for building a virtual Arctic spatial data infrastructure. *Computers & Geosciences*, *37*, 1752–1762. doi:10.1016/j.cageo.2011.06.024

Li, W., Yang, C., & Raskin, R. (2008). A semantic enhanced search for spatial web portals. In *Proceedings of the Technical Report of Semantic Scientific Knowledge Integration of AAAI Spring Symposium* (pp. 47-50).

Li, W., Yang, C., & Yang, C. (2010). An active crawler for discovering geospatial web services and their distribution pattern. *International Journal of Geographical Information Science*, *24*(8), 1127–1147. doi:10.1080/13658810903514172

Li, Z., Yang, C., Wu, H., Li, W., & Miao, L. (2010). An optimized framework for seamlessly integrating OGC Web Services to support geospatial sciences. *International Journal of Geographical Information Science*, *25*(4).

Liu, K., Yang, C., Li, W., Li, Z., Wu, H., Rezgui, A., & Xia, J. (2011, June 24-26). The GEOSS clearinghouse high performance search engine. In *Proceedings of the 19th International Conference on Geoinformatics*, Shanghai, China.

Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, ▪▪▪, 2.

Multitier architecture. (n.d.). In *Wikipedia*. Retrieved September 13, 2011, from http://en.wikipedia.org/wiki/Multitier_architecture

NASA. (n.d.). *NASA world wind*. Retrieved September 28, 2011, from http://worldwind.arc.nasa.gov/features.html

Nebert, D. (2007). OpenGIS catalog services specification, Version 2.0.2. *07-006r1_OpenGIS_Catalogue_Services_Specification_V2.0.2*, Open GIS Consortium Inc. Retrieved September 28, 2011, from http://www.opengeospatial.org/standards/cat

NSF. (2011). *Dear colleague letter: The "Earth Cube" - Towards a national data infrastructure for Earth System Science*. Retrieved October 19, 2011, from http://www.nsf.gov/pubs/2011/nsf11065/nsf11065.jsp?org=NSF

OpenLayers. (n.d.). Retrieved September 28, 2011, from http://openlayers.org/

OpenSearch. (n.d.). Retrieved September 28, 2011, from http://www.opensearch.org/Home

OWL web ontology language use cases and requirements. (2004). Retrieved September 28, 2011, from http://www.w3.org/TR/webont-req/

OWLGrEd. (n.d.). Retrieved September 28, 2011, from http://owlgred.lumii.lv

Paradise, J., & Beard, K. (1994). Visualization of spatial data quality for the decision-maker: A data quality filter. *Journal of the Urban and Regional Information Systems Association*, *6*, 25–34.

POST(HTTP). (n.d.). In *Wikipedia*. Retrieved September 14, 2011, from http://en.wikipedia.org/wiki/POST_(HTTP)

Precision and recall. (n.d.). In *Wikipedia*. Retrieved September 1, 2011, from http://en.wikipedia.org/wiki/Precision_and_recall

Protégé (Software). (n.d.). In *Wikipedia*. Retrieved October 13, 2011, from: http://en.wikipedia.org/wiki/Prot%C3%A9g%C3%A9_%28software%29

Rada, R., Mili, H., Bicknell, E., & Bletner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(1), 17–30. doi:10.1109/21.24528

Ramachandran, R., Graves, S., Conover, H., & Moe, K. (2004). Earth science markup language (ESML): A solution for scientific data-application interoperability problems. *Computers & Geosciences*, *30*(1), 117–124. doi:10.1016/j.cageo.2003.08.010

Ramirez, A. (2000). *Three-tier architecture*. Retrieved October 13, 2011, from http://www.linuxjournal.com/article/3508

Raskin, R., & Pan, M. (2005). Knowledge representation in the semantic Web for Earth and environmental terminology (SWEET). *Computers & Geosciences*, *31*, 1119–1125. doi:10.1016/j.cageo.2004.12.004

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of*, *IJCAI-95*, 448–453.

Rips, J., Shoben, J., & Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. doi:10.1016/S0022-5371(73)80056-8

Semantic search. (n.d.). In *Wikipedia*. Retrieved October 1, 2011, from http://en.wikipedia.org/wiki/Semantic_search

SPARQL query language for RDF. (2008). Retrieved October 13, 2011, from http://www.w3.org/TR/rdf-sparql-query/

SWEET ontologies. (n.d.). Retrieved Septerber 15, 2011, from http://sweet.jpl.nasa.gov/ontology/

Tessaris, S., et al. (Eds.). (2009). Reasoning Web 2009 (LNCS 5689, pp. 1-39).

Web feature service. (n.d.). Retrieved October 13, 2011, from http://www.opengeospatial.org/standards/wfs

Web map service. (n.d.). Retrieved October 9, 2011, from http://www.opengeospatial.org/standards/wms

Yang, C., Dong, M., & Fotouhi, F. (2005, November 6-11). Semantic feedback for interactive image retrieval. *Proceedings of the 13th Annual ACM International Conference on Multimedia*, Hilton, Singapore.

Yang, C., Goodchild, M. F., Huang, Q., Nebert, D., Raskin, R., & Xu, Y. et al. (2011). Spatial cloud computing: How can the geospatial sciences use and help shapecloud computing? *International Journal of Digital Earth*, *4*(4), 305–329. doi:10.1080/17538947.2011.587547

Yang, C., Li, W., Xie, J., & Zhou, B. (2008). Distributed geospatial information processing: Sharing earth science information to support digital Earth. *International Journal of Digital Earth*, *1*(3), 259–278. doi:10.1080/17538940802037954

Yang, C., & Raskin, R. (2010). Geospatial cyberinfrastructue (GCI): Past, present, and future. *Computers, Environment and Urban Systems*.

Yang, P., Evans, J., Cole, M., Alameh, N., Marley, S., & Bambacus, M. (2007). The emerging concepts and applications of the spatial web portal. *Photogrammetric Engineering and Remote Sensing*, *73*(6), 691–698. doi:10.14358/PERS.73.6.691

Yu, L. (2010). *A developer's guide to the semantic web*. Atlanta, GA: Springer Heidelberg.

Zhang, C., Li, W., & Zhao, T. (2007). Geospatial data sharing based on geospatial semantic Web technologies. *Journal of Spatial Science*, *52*(2), 35–49. doi: 10.1080/14498596.2007.9635121

Zhang, C., Zhao, T., & Li, W. (2010a). The framework of a geospatial semantic Web based spatial decision support system for digital Earth. *International Journal of Digital Earth*, *3*(2), 111–134. doi:10.1080/17538940903373803

Zhang, C., Zhao, T., & Li, W. (2010b). Automatic search of geospatial features for disaster and emergency management. *International Journal of Applied Earth Observation and Geoinformation*, *12*(6), 409–418. doi:10.1016/j.jag.2010.05.004

Zhang, C., Zhao, T., Li, W., & Osleeb, J. (2010c). Towards logic-based geospatial feature discovery and integration using web feature service and geospatial semantic web. *International Journal of Geographical Information Science*, *24*(6), 903–923. doi:10.1080/13658810903240687

Zhao, T., Zhang, C., Wei, M., & Peng, Z.-R. (2008). Ontology-based geospatial data query and integration. *Lecture Notes in Computer Science LNCS5266*. *Geographic Information Science*, *5266*, 370–392. doi:10.1007/978-3-540-87473-7_24

Kai Liu is currently a graduate student in the Department of Geography and GeoInformation Sciences in the College of Science at George Mason University. Previously he was a visiting scholar at the Center of Intelligent Spatial Computing for Water/Energy Science (CISC), and worked for 4 years at Heilongjiang Bureau of Surveying and mapping in China. His previous education was acquired at Wuhan University, China, BA Geographic Information Science. His research focuses on Geospatial semantics and Geospatial metadata management.

Chaowei Yang is associate professor of GIScience in the GGS department of George Mason University. He is the founding director of CISC and the NSF I/UCRC for Spatiotemporal Thinking, Computing, and Applications. He published over 100 papers and edited 6 international journal special issues. He served as the president of Chinese Professionals in Geographic Information Science (CPGIS). He is the research committee chair of the university consortium of geographic information science. He has graduated over 10 graduate students, who serve as professors and GIScience professionals in the US and China.

Wenwen Li is assistant professor at GeoDa Center for Spatial Analysis and Computation, School of Geographical Sciences and Urban Planning at Arizona State University. Her research specializes in the fields of geographic information science and remote sensing, semantic interoperability, spatio-temporal data mining, spatial information retrieval and distributed geospatial information processing.

Zhipeng Gui is a PostDoc in Department of Geography and GeoInformation Sciences in the College of Science at George Mason University. He got his Ph.D. degree in Cartography & Geographic Information Engineering, Wuhan University, China, 2011. His research interests include geospatial web service composition, web service quality and its applications, web-based geospatial resource discovery and cloud computing.

Chen Xu is a PostDoc at the George Mason University. His research interests include volunteered geographic information, social media, GIScience, and cloud computing. Chen Xu has a Ph.D in Geographic Information Science from the Texas A&M University (2010).

Jizhe Xia is currently a PhD student in the Department of Geography and GeoInformation Sciences, George Mason University. His formal education was acquired at the South China Normal University, China. His research focuses on spatial Indexing, data discovery and spatiotemporal pattern mining.