

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, Tempe, AZ, 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 meta-data 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., 2010), 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/geonet-work>) 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

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/using-semantic-search-and-knowledge-reasoning-to-improve-the-discovery-of-earth-science-records/111100

Related Content

An Examination of Job Titles Used for GIScience Professionals

Thomas A. Wikle (2012). *Geospatial Technologies and Advancing Geographic Decision Making: Issues and Trends* (pp. 68-81).

www.irma-international.org/chapter/examination-job-titles-used-giscience/63596

BIM Integrated Workflow Management and Monitoring System for Modular Buildings

Amar Seeam, Tianxin Zheng, Yong Lu, Asif Usmani and David Laurenson (2013). *International Journal of 3-D Information Modeling* (pp. 17-28).

www.irma-international.org/article/bim-integrated-workflow-management-monitoring/77814

Spatial-Temporal Analysis of Recent Air Passenger Flows

Changjoo Kim and Hyun Kim (2013). *International Journal of Applied Geospatial Research* (pp. 37-57).

www.irma-international.org/article/spatial-temporal-analysis-of-recent-air-passenger-flows/95193

Patient Confidentiality

Andrew Curtis and Michael Leitner (2006). *Geographic Information Systems and Public Health: Eliminating Perinatal Disparity* (pp. 224-244).

www.irma-international.org/chapter/patient-confidentiality/18856

Web Map Servers Data Formats

Yurai Núñez-Rodríguez (2009). *Handbook of Research on Geoinformatics* (pp. 82-88).

www.irma-international.org/chapter/web-map-servers-data-formats/20390