A Semantic Enhanced Model for Searching in Spatial Web Portals

Wenwen Li, Chaowei Yang
Joint Center for Intelligent Spatial Computing
College of Science, George Mason University
4400 University Dr.
Fairfax, VA U.S. 22030
{wli6,cyang3}@gmu.edu

Rob Raskin
Jet Propulsion Laboratory
4800 Oak Grove Dr
Pasadena, CA U.S. 91109
raskin@jpl.nasa.gov

Extended Abstract

Spatial Web Portal (SWP, Yang et al 2007) is used by the Earth science community in Earth science data sharing and exchanging. The SWP facilitates a large amount of geospatial resources, including text files, raw and post-processed data, and various geospatial web services. However, the popular utilizations also stand out problems of how to find needed data from a variety of geospatial resources and how to visualize the data from multi-perspectives. Currently, most search engines in SWPs are based on keyword matching or pure ontology based reasoning, which can not effectively 'understand' the meaning of user's queries, especially when a user has limited Earth science knowledge. For example, after California fire, people may ask more about "What's the air pollution caused by Southern California Fire in 2007?" (We define it as query Q1), but usually they do not know what pollutant may cause air pollution, or what factor is used in Air Quality (AQ) community to measure how bad the air pollution is. The above information is domain knowledge that is hidden behind the query and can not be inferred directly by keyword based search engines. Thus how to provide a 'specialized' answer through 'unspecialized' queries becomes an urgent challenge, which is also known as "Intelligent Question Answering" problem.

To solve the problem, this extended abstract reports our efforts on utilizing semantic web techniques (Berners-Lee, Hendler and Lassila, 2001) to enhance traditional search engines by: 1) building a semantic-based information model, which is a network of concepts with explicit relationships for implementing knowledge reasoning. With this model, a generic query can be explained and inferred to specific information needed. 2) Combining search tools provided within and cross SWPs to combine multi-layers of heterogeneous data sources together, results will include both web pages for presenting text information and spatial maps generated from remote web services for interactive search.

The Semantic Model

Our model is based on SWEET (Semantic Web for Earth and Environmental Terminology) ontology, where all the concepts are divided into different facets, including phenomena, property, substance, earth realm, to support reductionism (Raskin and Pan, 2005). As the most popular ontology model in Earth Science, SWEET provides an upper-level abstracted expression of the world. Based on formal description logic (DL), it can support terminology reasoning (T-box reasoning). However, only using this model may not be enough for answering query like Q1. First, "Southern California Fire in 2007" is an individual of "wildfire", which can not be inferred if there's no individual stored in the ontology. Second, formal query method is not able to retrieve the real resources (such as reports, documents or other data) even they've reached related node in the ontology model because this task is always partitioned as information retrieval (IR) task. Third, except for information provided in retrieved web pages, people may need more intuitive way, such as real-time data or geospatial maps which have sufficient resources in SWPs. So here we try to build an extended model based on SWEET combining both semantic web and IR techniques to understand and answer users' queries better in the following steps:

(1) Extracting information from queries and using them as input of our semantic model. Here we'll not focus on converting the natural language descript query to explicit DL expressions, but to extract keywords and do a basic analysis of them. For example, certain spatial queries may include place name and time such as 'California' and '2007' in Q1, so we'll measure the occurrence of place name and time factors. Meanwhile, we'll extract other keywords from the query, too.

(2) Extending SWEET ontology by adding more individuals as well as roles to support individual reasoning (A-box) so that to expand user's query. Figure 1 shows an example of the ontology fragment. It's visualized as a labeled graph: nodes in circles denote to terminologies and those in rectangles denote to individuals. Among which, the

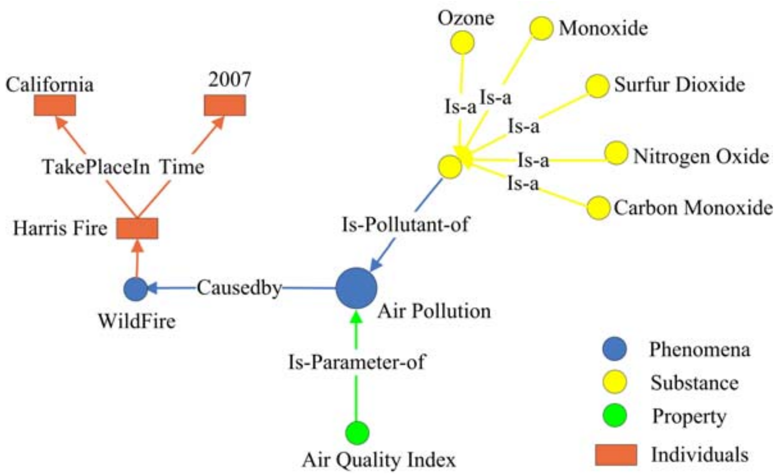

Figure 1. A graph of ontology fragment pertaining air pollution.

nodes of same color means they're in the same facet of ontology model. Different nodes are connected by certain roles labeled on the arrow. In this abstract, we'll take the fragment of ontology shown in Figure1 as a case study instead of talking about the whole A-box and T-box of the whole ontology model. By this model, query Q1 can be expanded to DL concept Q1a, Q1b and Q1c by formal query method (Horrocks and Tessaris, 2002):

Q1a: Fire. Name$\cap\exists$ cause.AirPollution$\cap\forall$ takePlacein.CA $\forall$ Time.2007
Q1b:
Pollutant$\cap\forall$ isPollutantOf.(AirPollution$\cap\forall$ causedBy.Fire)
Q1c:
Parameter$\cap\exists$ isParameterOf.AirPollution

Where, Q1a aims to find the names<n1, n2…nk> of all the fires that satisfy the restrictions. Q1b and Q1c are formed by checking all the roles that are connected with 'AirPollution' to get the pollutants<Po1, Po2…, Pom> and parameter<Pa> that's used for measuring air pollution separately. So, "Harris Fire" can be inferred from Q1a, and ("Ozone", "Monoxide", "Surfur Dioxide", "Nitrogen Oxide", "Carbon Monoxide") can be inferred from Q1b, and "Air Quality Index" can be inferred from Q1c (reference Figure 1).

(3) Coupling keyword groups together in an appropriate manner. As we can get more domain specific information (in the format of keyword) from the extended queries by applying DL A-box and T-box reasoning algorithm, so after coupling them together, such as <"Harris Fire", "Ozone", "Air Quality Index"> or <"Harris Fire", "Carbon Oxide", "Air Quality Index">, the chance of getting more accurate information must be increased.

(4) Dispatching the coupled keywords into search interface. Usually, there are two kinds of search tools provided within the SWP: one is IR based search tools to search for web documents. Also, considering the characteristic of spatial issues, most of the SWPs provides spatial web service search engine, such as in ESIP's Earth Information Exchange Portal (EIE) (Yang et al., 2007) and NASA's Earth Science Gateway portal (ESG) (Bambacus and Evans, 2005), they both provides basic search and service search functionalities. Thus related information found through them can be visualized into two types: text and generated geospatial map (As in Figure 2, "in SWP" box shows).

Integrating with Heterogeneous Data Sources cross SWPs

Resources from single SWP may not have sufficient information for certain application, so we try to build connections cross multi-SWPs to extend searching scales. As Figure 2 shows, after reasoning from semantic model (Figure 2, middle part), the system will redirect the keywords to GCMD and NCDC's web portal to retrieve more related web documents through the Application Programming Interfaces (APIs) (Figure 2, left part). Meanwhile, for discovering more geospatial map, we rely on OGC's Web Catalog Service (CSW) provided by Geospatial One Stop (GOS) portal and ESG portal for searching related map layers based on the combined keywords. Notice that, for CSW search, we can not send the keywords directly to the interface the same as web document search, but the keywords should be encapsulated as filters in a XML-based or KVP (Key Value Pairs)-based query according to OGC CSW specification (Nebert and Whiteside, 2005). Then, returned map layers and those got from local SWP will be integrated into a single layer and visualized in the query interface (Figure 2, right part). In this way, geospatial resources located in distributed environment could be seamlessly integrated together and in the meantime serve as a solid information base to support Earth Science information discovery.

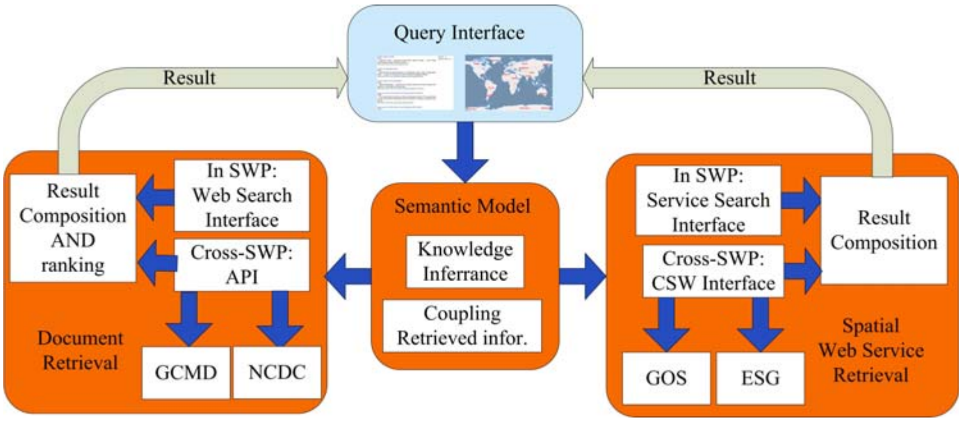

Figure 2. Integral Infrastructure for knowledge discovering.

Conclusion and Future Work

The research addresses intelligent question answering in Earth Science area. A semantic enabled model extending the search capabilities of existing methods in SWPs is able to answer more complex queries. In the latter part, we presents how to build connections cross popular SWPs to integrate and interoperate knowledge seamlessly.

Currently, we're developing knowledge discovering system for Air Quality to support effective decision making. In the future, we'll try to improve intelligent reasoning and developing systems that could support other societal benefit areas, such as Water Management, Carbon Management, Agriculture, Coastal Management, Homeland Security, Public Health, Invasive Species, and Disaster Management.

Acknowledgments

Research reported is supported by a NASA grant NNX07AD99G.

References

Berners-Lee, T.; Hendler, J. and Lassila, O. 2001. The Semantic Web. Scientific American 284, 34–43.

Bambacus, M.J. and Evans J. 2005. NASA's Earth-Sun System Gateway: An Open Standards-based Portal to Geospatial Data and Services. In proceedings of IEEE International Symposium of Geoscience and Remote Sensing, 4228- 4231.

Horrocks I. and S. Tessaris, 2002. Querying the semantic web: A formal approach. In Proceedings of International Semantic Web Conference, 177-191, 2002.

Nebert, D., and Whiteside, A. 2005. Catalog Services, Version 2, OGC Implementation Specification. Available online at: http://portal.opengeospatial.org/files/?artifact_id=20555 , OGC (accessed 1 November, 2007).

Raskin R. and Pan M. 2003. Semantic Web for Earth and Environmental Terminology (SWEET), Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Florida, USA.

Yang C.; Evans J.; Cole M.; Alameh N.; Marley S. and Bambacus M., 2007. The Emerging Concepts and Applications of the Spatial Web Portal, Photogrammetry Engineering &Remote Sensing, 73(6):691-698.

Yang, C.; LI, W.; Xiao, D.; Raskin, R. and Bambacus, M. 2007. Earth Information Exchange—Sharing Earth Science Information through Interoperable Approach and CyberInfrastructure. GeoInformatics'2007, Wuhan, China.