Data Integration for Digital GeoField Mapping

Erick Garcia[1,2,] Natalia Villanueva-Rosales [1,2,*]
1Department of Computer Science, [2]Cyber-ShARE Center of Excellence, [*]Faculty Mentor
El Paso, Texas, 79902

Introduction

This work is inspired in an on-going project using Augmented Reality (i.e., Digital GeoField Mapping) to enhance Geology data collection by providing additional information during the field trips. This project contributes to the challenge of integrating various disparate data sources using the vocabulary or the unifying ontology: The *GeoField Ontology*. Although there are huge amounts of data on the web available to the public, understanding and retrieving these data is challenging given that it is published in different formats and it may not have enough information to reuse the data. The data sources that we are used for this project include the National Oceanic and Atmospheric Administration's National Weather Service (NOAA; http://w1.weather.gov/xml/current_obs/) and United States Geological Survey (USGS; http://earthquake.usgs.gov/.).

Hypothesis

In this work we postulate that the annotation of Web data using formal vocabularies and the use of Web standards will streamline the integration of heterogeneous data on the web.

Methodology

The steps taken to complete the objectives of this project were the following:

1. Identifying the data sources relevant to the problem, i.e., the format of the data provided by the source and the data sharing service;
2. Transform and annotate data with formal vocabularies (i.e., ontology terms). We used the OWL API (http://owlapi.sourceforge.net/) to create an Ontology Populator. Using the Ontology Populator we created the *GeoWeatherReport* ontology with the weather XML data. The *GeoWeatherReport* ontology can be used for several applications given the generic descriptions of its classes. As an initial step, the data of the GeoWeatherReport was integrated with the GeoFieldOntology.
3. Validate the output data with respect to consistency with formal vocabularies and data loss in the transformation process. We validated our data with respect to consistency to the GeoWeatherReport.owl and GeoFieldOntology.owl ontologies.

To make the process of parsing the original data source and integrating it into our data source easier, two technologies (API's) where used:

- *Java DOM Parser:* The Document Object Model is an official recommendation of the World Wide Web Consortium (W3C). It defines an interface that enables programs to access and update the style, structure, and contents of XML documents. XML parsers that support the DOM implement that interface (https://docs.oracle.com/javase/tutorial/jaxp/dom/
- *OWL API*: A Java interface and implementation for the W3C Web Ontology Language (OWL), used to represent Semantic Web ontologies. The API is focused towards OWL 2 and offers an interface to inference engines and validation functionality.

Results

Through the use of the OWL API, we created an *Ontology Populator* in the project that can be easily extended to accommodate other sources of data using Web-based standards such as XML and RDF. Using the Ontology Populator we end up with a consistent and complete *GeoWeatherReport ontology* with data from the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service and the United States Geological Survey. This ontology contains all of the weather information in

Weather Reports, and its contents as data properties. The integrated data was consistent with the GeoWeatherReport ontology and the GeoFieldOntology ontology. By integrating these data, we can ask questions that involve reasoning, and answer questions that involve third-party data and domain knowledge provided by Geology experts.

Future Work

The future work includes the integration of additional data sources to evaluate the extensibility of the developed tools and integrating input of users, initially Geology students, that can validate the integrated data as well as inference drawn using the ontologies from the domain perspective.

Keywords: Ontology, Web Ontology Language (OWL), Extensive Markup Language (XML), Parsing, Document Object Model (DOM)