CYBER-ShARE Center of Excellence

Sharing Resources to Advance Research and Education through Cyber-infrastructure

Data Integration for Digital GeoTimescale Mapping

Erick Garcia[1,2], Natalia Villanueva-Rosales[1,2,*]

[1]Department of Computer Science, [2]Cyber-ShARE Center of Excellence, [*]Faculty Mentor

El Paso, Texas, 79902

Abstract

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 This work is inspired in an on-going project using Augmented Reality (i.e., Digital GeoTimescale Mapping) to enhance Geology data collection by providing additional information during the field trips. This project addresses the challenge of integrating disparate data sources using a controlled vocabulary from a unifying ontology: The GeoTimescale Ontology. The data sources that we are used for this project include the National Oceanic and Atmospheric Administration's National Weather Service [1] and United States Geological Survey [2].

Background

- An **Ontology** an ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse [3].
- **Web Ontology Language (OWL)** a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.[4].
- **Parsing** is the process of identifying a data structure, reading it and converting it to an usable format in memory [5].

Problem Statement

Even though there are huge amounts of data on the web, understanding and retrieving it can be challenging given that it is published in different formats and it may not have enough information to reuse the data

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 and Testing

1. **Identification** of the data sources relevant to the problem, i.e., the format of the data provided by the source and the data sharing service;
2. **Transformation and annotation** of the data with formal vocabularies (i.e., ontology terms). We used the OWL API[5] to create an Ontology Populator. Using the Ontology Populator we created the GeoWeatherReport ontology with the weather XML data. The GeoWeatherReport ontology [6] 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 GeoTimescale ontology[7]
3. **Validation** of 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 and GeoTimescale ontologies.

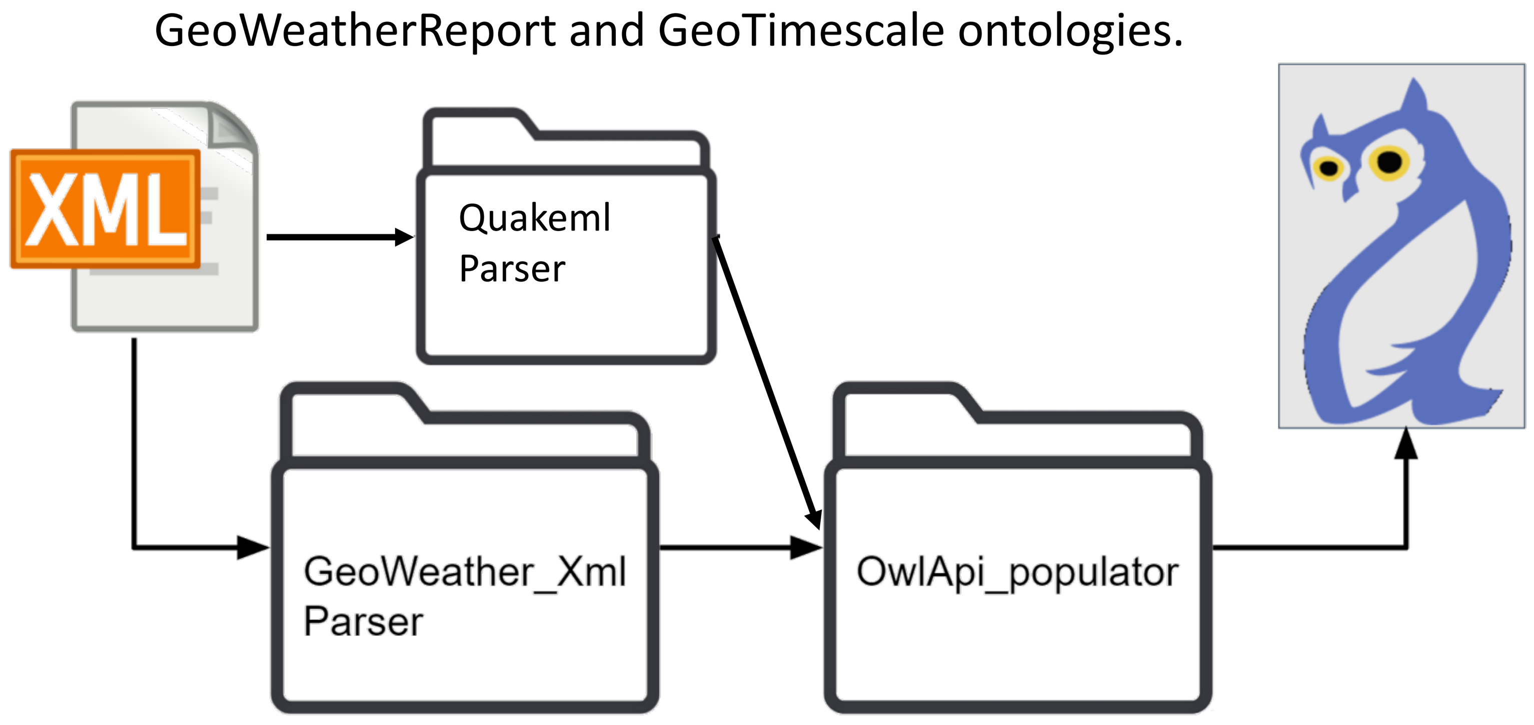

Figure 1. The data source Weather (XML) is processed by the GeoWeather_XML Parser, and finally it is converted into an ontology by the OwlApi_populator

References

[1] Service, N. N. W. (n.d.). XML data feeds - Current Conditions - NOAA's National Weather Service. Retrieved September 29, 2016, from http://w1.weather.gov/xml/current_obs/
[2] QuakeML Summary Format. (n.d.). Retrieved September 29, 2016, from https://earthquake.usgs.gov/earthquakes/feed/v1.0/quakeml.php
[3] Ontology (Computer Science) - definition in Encyclopedia of Database Systems. (n.d.). Retrieved September 29, 2016, from http://web.dfc.unibo.it/buzzetti/IUcorso2007-08/mdidattici/ontology-definition-2007.htm
[4] OWL - Semantic Web Standards. (n.d.). Retrieved September 29, 2016, from https://www.w3.org/OWL/
[5] DOM Parsing, Retrieved September 28, 2016 from https://www.w3.org/TR/DOM-Parsing/
[6] Cyber-ShARE Ontologies. (n.d.)., GeoTimescale Ontology, Retrieved September 29, 2016, from http://ontology.cybershare.utep.edu/geotimescale.owl
[7] Cyber-ShARE Ontologies. (n.d.)., GeoWeatherReport Ontology, Retrieved September 29, 2016, from http://ontology.cybershare.utep.edu/geoweatherreport.owl
[8] Lesson: Document Object Model (The JavaTM Tutorials > Java API for XML Processing (JAXP)). (n.d.). Retrieved September 29, 2016, from https://docs.oracle.com/javase/tutorial/jaxp/dom/
[9] OWL API. (n.d.). Retrieved September 29, 2016, from https://sourceforge.net/projects/owlapi/
[10] egarcia87 / Digital-ROX_code. (n.d.). Retrieved September 29, 2016, from https://bitbucket.org/egarcia87/digital-rox_code

Technologies Used

- Java DOM Parser: The Document Object Model is an official recommendation of the World Wide Web Consortium (W3C) [3]. 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 [8].
- OWL API: A Java interface and implementation for the W3C Web Ontology Language (OWL) [4], used to represent Semantic Web ontologies. The API is focused towards OWL 2 and offers an interface to inference engines and validation functionality [9]

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 [1] and the United States Geological Survey [2]. 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 GeoTimescale 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 [10].

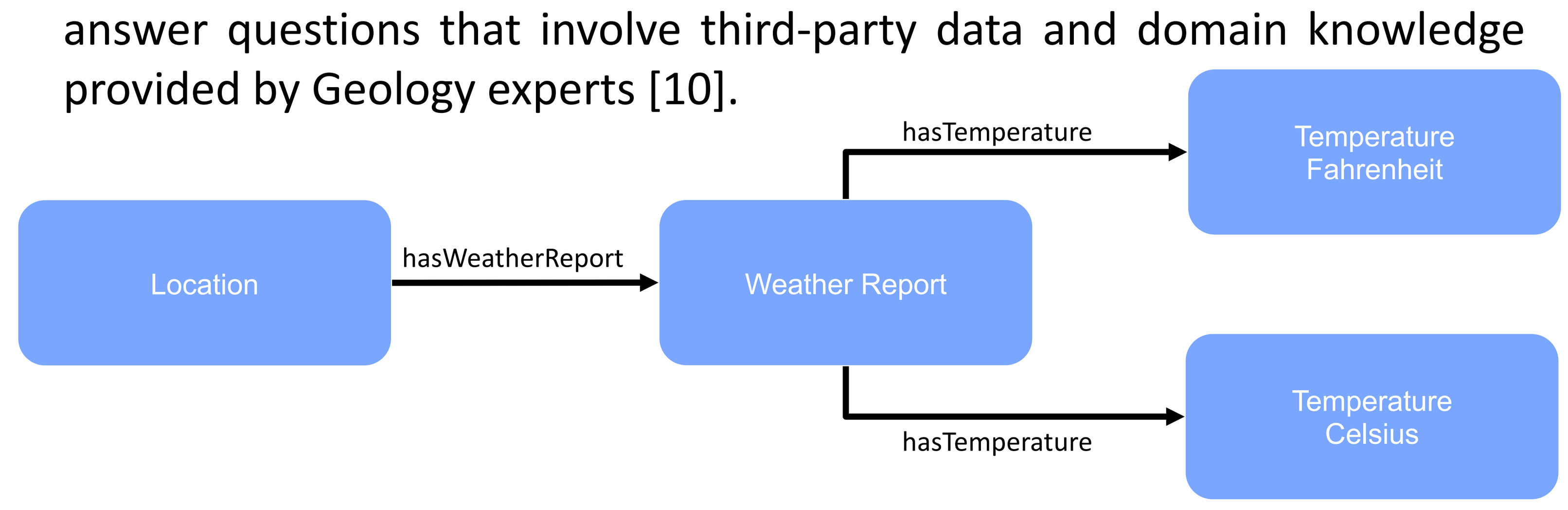

Figure 2. The figure shows the structure of the GeoWeatherReport Ontology and how each weather report is identified by location when imported into the main GeoTimescale Ontology

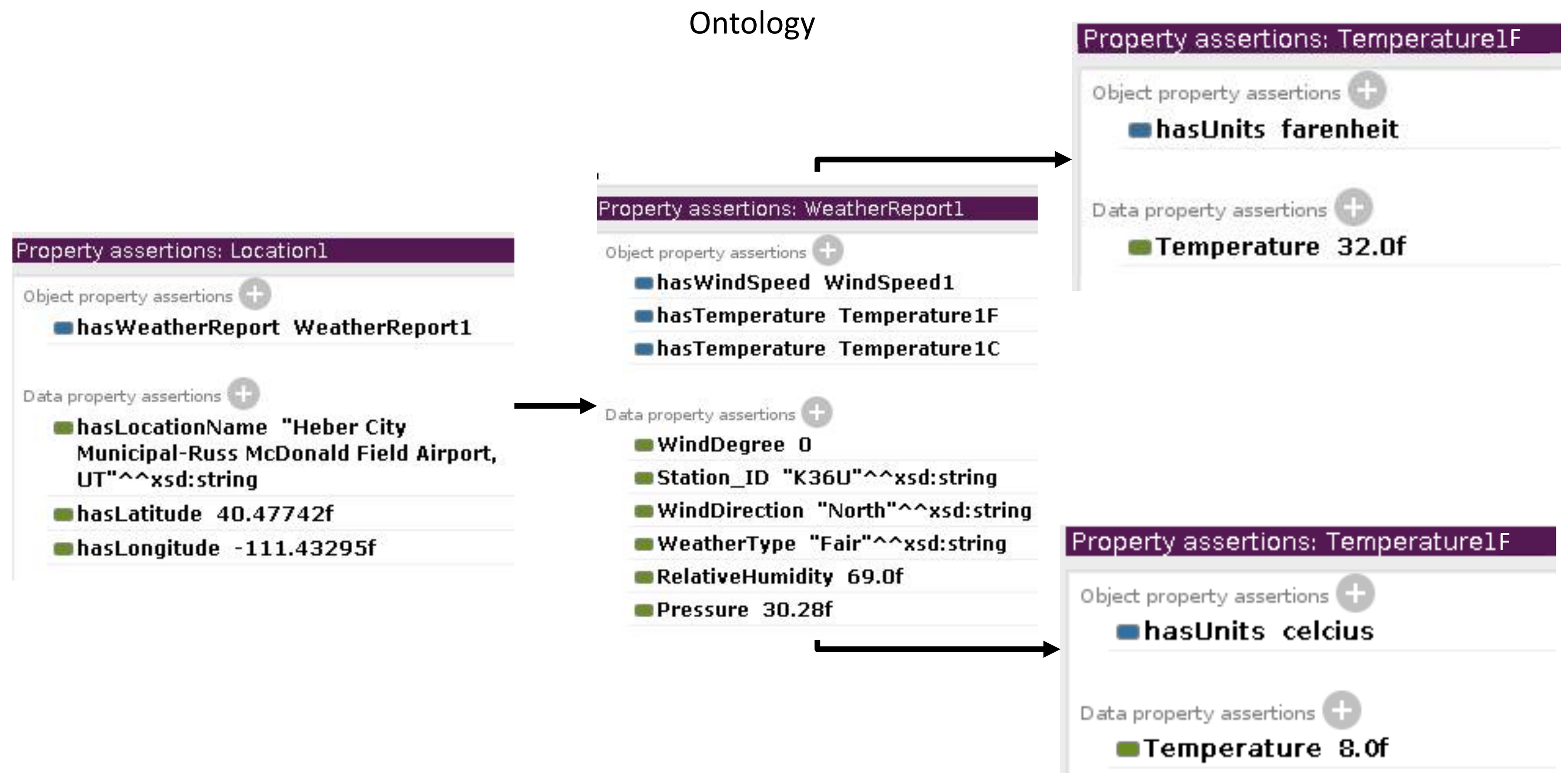

Figure 3. Individuals from the ontology with real data from the parsed XML's

Conclusion

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.

Acknowledgements

This project contributed to the Digital GeoTimescale Mapping led by Perry Houser, was integrated with a question answering module created by Richard Samples and Georgia Almodovar and used XML data retrieved by Smriti Rajkarnikar Tamrakar – all of them students at UTEP.

NSF CAHSI

This work used resources from the Cyber-ShARE Center of Excellence and the Computing Alliance of Hispanic Serving Institutions (CAHSI), which are supported by National Science Foundation, grant numbers HRD-0734825 and CNS-1042341 respectively. Special thanks to the iLink research group members for their support in this work.

UTEP Cyber-ShARE iLink Linking knowledge across disciplines, data and models