Semantically Enabled Temporal Reasoning in a Virtual Observatory

Patrick West, Eric Rozell, Stephan Zednik, Peter Fox, Deborah McGuinness
{westp,rozele,zednis}@rpi.edu {pfox,dlm}@cs.rpi.edu

Rensselaer Polytechnic Institute, Tetherless World Constellation

Abstract. The Virtual Solar-Terrestrial Observatory (VSTO) is a distributed, scalable education and research environment for searching, integrating, and analyzing observational, experimental and model databases in fields of solar, solar-terrestrial and space physics.
Our work on VSTO required us to create a formal, machine understandable representation for concepts, relations and attributes of physical quantities, including concepts to represent scientific instruments and their operating modes, parameters measured by these instruments, and measured from a particular location during a particular time period. The end-user is allowed to search for data based on measured parameters of interest, instruments of interest, geophysical constraints and temporal constraints. To fulfill this need we developed a set of formal encodings of the knowledge in the OWL-DL format.
In the current implementation of VSTO (www.vsto.org), however, time coverage information is retrieved as needed from a relational database, not from the ontology. Consequently, the implementation is only able to perform temporal reasoning of the subsets of time intervals returned. This choice was made because it addressed our concerns, given the current level of the tools available at the time, of performance and scalability of creating millions of in-memory triples representing time coverage.
The ideal solution would be to generate instances to represent any temporal relationships within the data. This would provide a greater range of temporal reasoning and search capabilities for the user.
This paper will present information on how we are working to take the next step in this solution and to answer the questions of extending the expressiveness of the ontology to include more complex temporal reasoning; how this will impact the tools used to implement the ontology, how this will impact the implementation of the underlying data portal, and how this will impact the users of the system, the domain scientists, and their ability to effectively search for the data needed in a timely manner.

1 Introduction

The Virtual Solar-Terrestrial Observatory (VSTO) is a multidisciplinary virtual observatory providing what appears to the user to be one integrated local resource for heterogeneous distributed scientific data, theoretical models and analysis programs. The integration of these heterogeneous systems is accomplished through the use of semantically enabled technologies, creating encodings of the semantics of the data to facilitate interoperability. The formal encodings of the knowledge was developed using OWL-DL, the Description Logics dialect of the Web Ontology Language.

The scientific domains included within the VSTO are solar physics, space physics and solar-terrestrial physics. Major communities include those interested in solar images from the Mauna Loa Solar Observatory (MLSO), and the NSF-funded Coupled Energetics and Dynamics of Atmospheric Regions (CEDAR).

The current implementation of the VSTO data portal can be found at http://vsto.rpi.edu. We use CMapTools to create the ontology model; Protege to create the actual ontology, which can be found at http://escience.rpi.edu/schemas/vsto_all.owl, and to generate Java code; the Java code being used as part of a Java Servlet application running under Apache Tomcat. The triples are loaded into memory using Jena and we utilize the functionality of Pellet for reasoning.

2 The Problem

The problem that we are attempting to resolve is the fact that our current ontology does not utilize extensive temporal information. All temporal information is retrieved via a REST-ful (Representational State Transfer)

Proceedings of OWL: Experiences and Directions 2009 (OWLED 2009),
Rinke Hoekstra and Peter F. Patel-Schneider, editors. http://www.webont.org/owled/2009

service call, gathering information based on the current selections that the user has made. For example, the user selects a particular CEDAR instrument they are interested in; a REST-ful service call is made in order to retrieve all available years in which the instrument has data. Once a year is selected, the list of available months is retrieved via the REST-ful service, and then the day of the month. The use of this metadata service call is modeled in the VSTO ontology (Fig. 1.,) however, the temporal information itself is not. This restricts our ability to perform temporal reasoning and restricts the users ability to search using more complex temporal constraints.

3 Initial Solution

The initial solution (2005) to implementing VSTO was based on a number of factors. First off, we were uncertain how well the tools would be able to handle the amount of time instances that we would be providing, estimated to be around 16 million. Would our chosen triple-store, Jena, be able to handle that many instances? Would the semantic reasoner, Pellet, be able to query against that many instances? What would performance be like? We were also unsure as to how this would be modeled in the ontology given these possible limitations. For this reason, we decided to model temporal concepts in the same way that the scientists were already comfortable selecting date/time.

Another reason is that we felt the first implementation of a semantically enabled data search/browse interface for scientists should be as close to what they were used to as possible. We wanted an introduction to semantically enabled scientic tools that wasn't too extensive for the scientists to take in for a first experience. Scientists become used to searching and accessing data from particular places using particular tools in particular ways. They become used to the way the data is organized and presented, and the data products provided. In informal discussions with users, signicant changes in the user interface was not going to be accepted by the specific communities. With this in mind we first wanted to simply replicate the functionality of the current CEDAR and MLSO data portal while adding semantic content at the same time.

In the original CEDAR portal, the scientist was guided through a search for their data, starting with an instrument or a single date/duration specification. After selecting the instrument the user was directed to the next step in the guided search, and then the next, until, finally, they were directed to their data.

To introduce scientists to semantically enabled data search, we introduced them to the idea of viewing the content behind their selection. For example, when given the selection of an instrument, the user can see descriptive information about this instrument, the type of instrument (an instrument class hierarchy), and the user has the ability to filter their search for an instrument by scientific domain (CEDAR or MLSO) or the type of instrument (instrument class). Also, we wanted to add semantic reasoning behind the scenes. For example, if a scientist selects a parameter with an associated coordinate parameter, then the parameter could be plotted as, say, a time-series plot. So, depending on the parameters selected, a different set of data products could be presented to the user.

With respect to the date/time selection, we kept this as similar as possible for the user to their original system. The user is provided with the ability to select a starting date, and the number of days they are interested in. Rather than modeling this in the ontology we modeled metadata service classes that know how to go out and retrieve this information (Fig. 1.)

Now that we have introduced the scientists to the idea of semantically enabled data search and the use of semantically enabled tools, we can expand the search capabilities to include faceted search, more semantic reasoning, and more expansive temporal constraints.

4 Ideal Solution

In the ideal situation, the VSTO ontology would contain all temporal information necessary to support the selection of any time interval the end-user may consider. Every record in the case of the CEDAR domain and every image in the case of the MLSO domain would be associated with a time instance or coverage interval (Fig. 2.) The ontology would utilize a reasoner, such as Pellet, along with the time instance data and data product coverage in various types of use case driven temporal reasoning. All of this reasoning will be supported by the addition of the OWL-Time ontology [3] to the VSTO ontology.

The current use cases for the VSTO projects can be generalized as such, retrieve data (from appropriate collections) subject to (stated and implicit) constraints and create a representation of the data in a manner appropriate for the data and for the end-user [2]. The two use cases that will be supported by the addition of temporal reasoning capabilities follow this general form, they are: retrieve data within a convex time interval and present it in a manner appropriate for the data and the end-user, and, retrieve data within a non-convex time interval and present it in a manner appropriate for the data and the end-user. A time interval is convex if it can be determined by its end points, for example March 1, 2009 through March 31, 2009. A non-convex time interval would be composed of more than one convex time interval, for example, the first Wednesday of every month, where each Wednesday is a convex time interval. [1].

For convex time intervals, the time selection portion of the data portal will remain very similar to its current state. The user will be provided a guided workflow and all possible selections (i.e. instrument, time interval, and parameter) will draw from instances within the ontology. Here the ontology, along with the Pellet reasoner, will essentially replace the metadata service of the initial implementation. The data portal will then take the users constraint selection and generate the URL that will provide them with the data they are looking for.

The more important part of this extension of reasoning, however, is the ability to select non-convex time intervals, as this is something the data portal does not currently support. It is likely that the first step towards providing the user with fully flexible temporal reasoning will be providing support for regular non-convex intervals. For a time interval to be regular non-convex, it must be a non-convex interval consisting of a series of identical convex intervals [1]. For regular non-convex intervals, the user would first select the smaller convex interval, which could be anything from a specific day within a week, or day number within a month, to an entire week or month. Next, they would select an overall time interval and pattern that would represent the regular non-convex interval from which data will be pulled. An example of a use case that this type of reasoning will support is: Retrieve data for all parameters measured by the Millstone Hill Fabry-Perot on each Wednesday of March 2001. The way that data products are presented to the user will have to be much different than that for convex time intervals.

5 Current Work

We are currently making progress towards adding flexible temporal reasoning to the VSTO data portal. The addition of temporal concepts to the VSTO ontology has been achieved by importing the OWL-Time ontology. With OWL-Time, we can represent relationships between instants and intervals, duration information, as well as date-time information [3]. The VSTO ontology has been updated with additional triples linking instruments to the parameters they measure. Property relations between time instances and the instruments and parameters available were created with the addition of just a few new object properties.

Some of the current objectives include generating time instance data and modifying the portal code to support the convex time interval use case. Changes to the web portal code will make it possible to utilize the new time instances. The VSTO data portal has also been moved to a newer, faster machine with more memory and more disk space. The Pellet Reasoner has been upgraded as has the Jena triple store. We will also be trying out different triple stores, ones that store triples in memory as well as those that store them persistently. Once we have completed our changes, information regarding the ability to scale such a large number of instances, and the performance when compared to the use of REST-ful service calls, will be revealed.

6 Acknowledgements

The VSTO project is funded by the National Science Foundation, Office of Cyber Infrastructure under the SEI+II program, grant number 0431153.

7 Figures

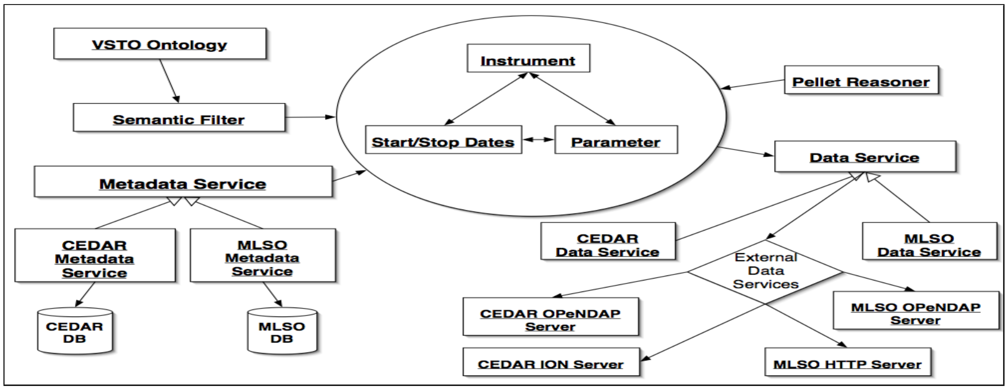

Fig. 1. Current modeling of metadata service calls in VSTO ontology. Individuals of the Metadata Service class know how to execute requests to gather temporal information, instead of time instances being part of the ontology itself.

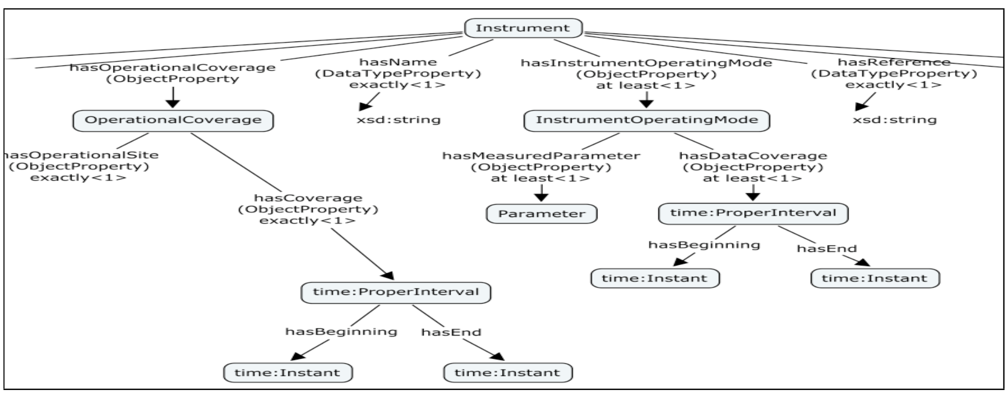

Fig. 2. New Potential Temporal Modeling. One use showing the convex time interval that an instrument is operating at a site. Another use showing the convex time interval that an instrument is operating in a particular mode.

References

1. Q. Zhou and R. Fikes, *A Reusable Time Ontology*. KSL-00-01, Stanford University, 2000.
2. Fox, P, D L McGuinness, L Cinquini, P West, J Garcia, J Benedict, and D Middleton, 2009, *Ontology-supported Scientific Data Frameworks: The Virtual Solar-Terrestrial Observatory Experience.*, Computers and Geosciences. Vol. 35, Issue 4, pp 724-738.
3. J. Hobbs and F. Pan, *Time Ontology in OWL*, W3C Working Draft, 27 Sept. 2006, http://www.w3.org/TR/owl-time/.
4. McGuinness, D.L., Fox, P., Cinquini, L., West, P., Benedict, J., Darnell, J.A,, Garcia, J., Middleton, D., 2006, *Experiences using Semantic Technologies in Virtual Observatories*', OWL Experiences and Directions,

CEUR Workshop Proceedings, vol. 216, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-216/submission_14.pdf

5. D. McGuinness, P. Fox, L. Cinquini, P. West, J. Benedict, and J. Garcia, 2007, *Current and future uses of OWL for scientific data frameworks: successes and limitations'*, OWL Experiences and Directions 2007, to be published in CEUR Workshop Proceedings, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-258/paper32.pdf