Semantic Model for Solar Data Organization and Sharing: CRAAM-SVO

Israel dos Santos[b,c 1], Nizam Omar[b,c] and Adriana Benetti Marques Valio[a,c]

[a]Center for Radio Astronomy and Astrophysics Mackenzie, Engineering School, Mackenzie Presbyterian University, São Paulo, SP, Brazil

[b]Electrical Engineering and Computation Graduate Program (PPGEEC), Mackenzie Presbyterian University, São Paulo, SP, Brazil

[c]Faculty of Computation and Informatics (FCI), Mackenzie Presbyterian University, São Paulo, SP, Brazil

Received on September 5, 2016 / accepted on October, 2016

Abstract

Solar flares are the most energetic phenomena in the Solar System, and last from seconds to hours. The Mackenzie Center for Radio Astronomy and Astrophysics (CRAAM) monitors solar activity at different radio wavelengths generating Total Power light curves and maps of the Sun, for Internet sharing. However, the uncontrolled growth in two decades of the global system of computer networks has shown large amount of data and sources with various and ambiguous contents. The Semantic Web allows the organization of this content through the merger of ontologies, standards, and technologies to provide the organization and intelligent interaction of content for maximum possible shared information about a resource. Thus, the data sharing from a virtual observatory, according to the International Virtual Observatory Alliance (IVOA), must have an ontology, which is a formal and explicit specification of a shared conceptualization, for the representation of knowledge in the field of Astronomy and open sharing of data in XML (eXtensible Markup Language). Therefore, the CRAAM Solar Virtual Observatory was developed in Java Enterprise Edition architecture with SOAP web services (Simple Object Access Protocol) integrated into the hybrid database Virtuoso. This database provides object-relational storage making use of ontology to support reading, inferences, and manipulation of data in a semantic model by RDF (Resource Description Framework). The result is a metadata catalog that can be accessed through interactive data search and SPARQL queries engine.

Keywords: solar virtual observatory, ontology, semantic web.

[1]E-mail Corresponding Author: israel.santos@mackenzie.br

1. Introduction

Solar flares occurring on the atmosphere of the Sun are related to underlying changes in magnetic fields. Large amount of energy is released during the explosions in the form of radiation, covering the entire electromagnetic spectrum, from radio waves to X-rays and gamma rays. Associated with solar activity are coronal mass ejections, which occasionally come toward Earth.

The experiments of the Italian mathematician and physicist Galileo Galilei (1564-1642) allowed humanity to refute the empirical conception that dominated science since the time of Aristotle, in the 4th century BC. After that, humanity has migrated from empirical science, conducted on the basis of descriptions of natural phenomena, to the validation of theories. This principle has allowed science to develop and start experimentation, monitoring, and collection of data, generating new knowledge. With the advent and development of technology, the quantity and quality of data has grown considerably, at a rate of 2 to 4 Petabytes per year.

The Information and Communication Technologies (ICT) have evolved exponentially in the last decade. The democratization and universal access to knowledge and information allowed researchers and simple users acess to large amount of data from different sources and repositories. Astronomy benefited from these advances [1], i.e, the international astronomical community has accumulated a lot of data from Earth and space, covering virtually the entire electromagnetic spectrum. However, data, metadata, interfaces and accessibility are heterogeneous and distributed in a range of custom databases [1].

In this sense, recent advances in the area of infrastructure, with the advent of fiber optics, cloud computing, virtualization, web services platform, database expansion, among others, contributed to the creation of a new paradigm called e-Science, proposed by Jim Gray [2]. He presented the different aspects necessary for integration of information technology and other fields of knowledge. Carvalho et al. discoursed about the fundamentals of e-Science for building Virtual Observatories, to address the scientific and technological challenges inherent to complex sets of existing data [1].

This article goal is to present the specifications for the construction of the Virtual Solar Observatory of the Mackenzie Center for Radio Astronomy and Astrophysics, using the precepts of the e-Science, aimed at expanding the access to solar data in organized and reliable mode, since currently these data are scattered on different platforms.

2. Semantic Web and Ontology

The Semantic Web is an extension of existing Web to attribute meaning to the content of the data, being composed by knowledge representation, ontologies and by agents. The Semantic Web is a framework that fosters understanding and Web content management, supplied in the form of text, sound, images or graphics [3].

No one knows for sure who first proposed the creation of a language on the World Wide Web that could be used to express the knowledge necessary to implement Web applications. However, even before most people knew the Web existence, there were several research groups in the middle of the 90s working on the idea of establishing conditions to improve the search and navigation job. In this decade, several researchers published algorithms that contained different approaches aimed at Web searching. Several projects included a set of concepts and relationships with the objective of developing a semantics to HTML pages and a semantic search engine. These projects are the current Web ontology languages and were mainly based on the Web needs and not on principles of artificial intelligence. At the same time, the W3C explored how some types of Web language could be defined with the aim to bring data from the Web, building, for this purpose, a language called Resource Description Format (RDF).

The Semantic Web is not an isolated Web but an extension of the existing Web. In this expansion, the information is given a well-defined meaning, to foster a cooperative work between computers and people.

The Semantic Web has emerged to solve mainly two particular problems existing in common Web [4]. One refers to limiting access to data since the documents are indexed and accessed by means of guided texts, which creates problems for ambiguous terms. Also, the usual paradigm is dominated by the return of the document that best fits the search performed and, finally, the unavailability of underlying data. Another common problem in the Web is the fact that it is seen as a large collection of static documents. Issues such as integration of information, data analysis, and making sense to the machines represent the only way followed by users, communities, and companies, to make the most of the information available on the Web.

Over the past decade, researchers from different areas worked on the idea of combining the semantics and the Web, resulting in three different visions of the Semantic Web. One is the Semantic Web seen as a layer of text, with the focus on the transformation of the Web of billions of textual documents in a structured and defined repository of semantic descriptions.

In this context, are considered the research that unites the Semantic Web with processing studies of natural language. A second view is the Semantic Web as a database, with the emphasis on the development of structures of knowledge as agents platforms. This kind of vision had great commercial success.

Finally, the Semantic Web view was developed as a platform for supporting agents that could answer questions for human users. The ontology-based data support semantic interoperability through meanings encoded in a manner that considers the semantic intersección. The development of the observatory described in this article follows the latter view.

The XML (Extensible Markup Language) was developed as a generic way to structure documents on the Web. This generalizes the HTML using defined issues to the user. Nevertheless, its flexibility reduces the chances of possible semantic interpretations with predefined questions from HTML.

The RDF (Resource Description Framework) is a simple data model that describes semantically the tools on the Web. RDF schema (RDFS) use basic statements in RDF and defines a simple ontology language. Specifically, it defines entities enabling class modeling, scope and domain restricted properties, and hierarchies of class and properties.

The Web Ontology Language (OWL) extends this vocabulary to a full spectrum of logical descriptions defined in RDF, called OWL Lite, OWL DL and OWL Full. Mechanisms are provided that allow properties definition, which can be classified into reverse, transitive, symmetry or functional. Properties can be used to set the instantiations to classes or class hierarchies and properties. OWL Lite is a significant logical description that develops efficient implementations for large data sets.

For the Observatory construction we used the OWL, considered as a formal language of the first order logic description family. This allows the representation of models made up of an organized vocabulary in hierarchy types such as local range of properties, treatment of disjoint classes, Boolean combination of classes, cardinality constraints, and special features of properties, among others.

The construction of this Observatory was founded on the Computer Ontology. In the literature of computational sciences there are several ontology concepts. For example, for Gruber, ontology is an explicit specification of a conceptualization [5]. To Guarino, ontology is the description of knowledge with a defined vocabulary within a domain [6] and to Borst, ontology is a formal specification of a shared conceptualization [7]. To present a more precise and formal definition of ontology, Guarino, Oberle and Staab seek to highlight the basic aspects that compose the word, highlighting the definitions of conceptualization, with an explicit specification and formal precision, emphasizing the importance of a construction for the common use [8].

> "Let **C** a conceptualization, and **L** a logical language with vocabulary **V** and ontological commitment **K**. An ontology $\mathbf{O}_k$ for **C** with vocabulary **V** and ontological commitment **K** is a logical theory consisting of a set of formulas of **L**, designed so that the set of its models approximates as well as possible the set of intended models of **L** according to **K**". (GUARINO; OBERLE; STAAB, 2009, p. 11)[8].

It is possible to observe that the definition constructed by these authors reveals the ontology as a logical theory, being considered ideal when the approximation is larger between its model and the intentional model.

Below, we present aspects of the construction of the Virtual Observatory called CRAAM-SVO, prepared using the methodology proposed by Guizzardi [9]. The methodology consists of six phases, as follows: purpose identification and specification of requirements, Capture, Drawing, Integration, Evaluation and Documentation. The life cycle takes place iteratively and the evaluation processes of the quality and the documents occur concomitantly to these phases.

3. Solar Virtual Observatory

The CRAAM monitors and collects data from various existing telescopes in the Southern Hemisphere and stores them in relational databases (PostgreSQL). These data remain available for processing and analysis in a timely fashion in its original format. For processing, the data receive initial conversion from ADC to temperatures, correction for atmospheric absorption, and calibration, to generate both light curves in Total Power as well as maps of the sun. Figure 1 shows the Sun Maps from the CRAAM-SVO portal.

The CRAAM-SVO architecture is composed of different layers of hardware (physical) and software (logical) to favor low maintenance, and seeking to organize the information so that search engines can take advantage and display the content in a variety of online applications. Figure 2 shows the result of the Web service call to generate light curves of the Submillimeter Solar Telescope (SST).

Considering the need for interoperability with other research centers and the semantics for metadata, the textit International Virtual Observatory Alliance (IVOA) posted on its website a standard ontology, from which it was possible to expand and represent the domain of CRAAM- SVO and generate the data source. The processed data is stored in Virtuoso Universal Server for conversion and dynamic display of relational data in RDF or XML.

Figure 1 - CRAAM-SVO Gateway

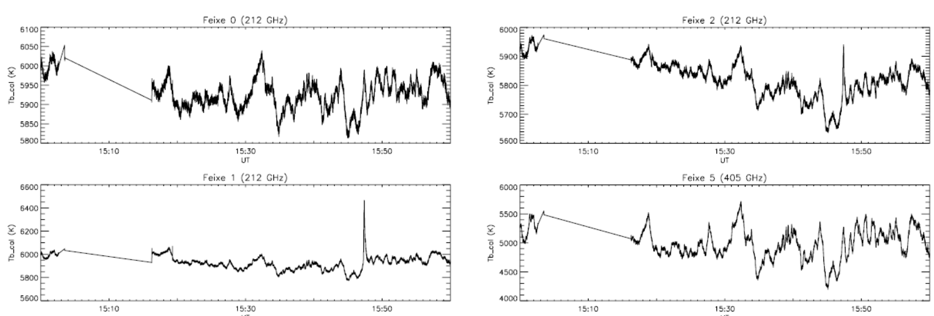

Figure 2 - Light curves of the Submillimeter Solar Telescope

Figure 3 shows the architecture of CRAAM-SVO system, which comprises an application server that supports requests from users and partners research centers. This component is developed in Java Enterprise Edition architecture using the Jena framework for integration to Virtuoso database through RDF / XML and API, to generate FITS files (Flexible Image Transport System) for data routing.

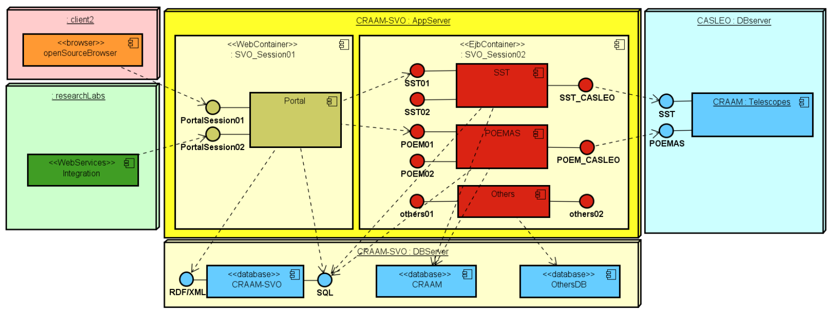

Figure 3 - Diagram of CRAAM-SVO Deployment

The files sent by the Observatories are directed to the respective data processing modules for each telescope. For example, for Submillimetre Solar Telescope, the data are forwarded to the SST component and divided into two processes. First, the data is persisted in the PostgreSQL database for historical and future recovery of the unprocessed data matrix. Second, the data are analyzed and persisted in Virtuoso server for reading, inferences and manipulations in OWL / RDF / XML and / or SQL. It is noteworthy that the SST module is hybrid because it uses the languages of Java and Python programming. Unit and load testing were performed, however usability evaluation tests will be done after making the system available to the scientific community .

4. Conclusion

To enable data sharing and solar information, a Virtual Observatory called CRAAM-SVO was developed based on the precepts of the e-Science. Guided by the Semantic Web and the ontology proposed by IVOA, the observatory from this study was developed to allow solar data from different instruments to be incorporated into a single portal. For this, each telescope has a component that allows the processing and storage of data, and the portal is the aggregator element of the system. Thus, the construction presented in this work allows the user to do research analysis and inferences starting from an organized environment, that is also reliable and shareable.

Acknowledgments. Israel dos Santos acknowledges financial support from CAPES and MackPesquisa.

References

[1] de Carvalho RR, Gal RR, de Campos Velho HF, Capelato HV, La Barbera F, Vasconcellos EC, Ruiz RS, Kohl-Moreira JL, Lopes PA, Soares-Santos M. The Brazilian Virtual Observatorya new paradigm for astronomy. Journal of Computational Interdisciplinary Sciences 1(3):187-206, 2010.

[2] Hey, A.J.G. and Tansley, S. and Tolle, K.M Jim Gray on eScience: A Transformed Scientific Method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11, 2007. In: The fourth paradigm: data-intensive scientific discovery, 2009, Microsoft research Redmond.

[3] Berners-Lee, T.; Hendler, J. Scientific publishing on the semantic web. Nature: 410, 1023-1024, 2001.

[4] Domingue, J., Fensel, D., & Hendler, J. A. (Eds.). Handbook of semantic web technologies, 2011, Springer Science & Business Media.

[5] Gruber, T. What is an ontology? janeiro 1993. WWW Site http://wwwksl.stanford.edu/kst/what-is-an-ontology.html (accessed on 2011-02-23).

[6] Guarino, N. Understanding, building and using ontologies. International Journal of Human Computer Studies, Academic Press: 46, 293-310, 1997.

[7] Borst, W. N. Construction of engineering ontologies for knowledge sharing and reuse. [S.l.]: Universiteit Twente, 1997. WWW Site http://doc.utwente.nl/17864/1/t0000004.pdf (accessed on 2011-03-02).

[8] Guarino, N.; Oberle, D.; Staab, S. What is an ontology. In: Handbook on Ontologies, 2009, Springer.

[9] Guizzardi, Giancarlo. Ontological foundations for structural conceptual models. CTIT, Centre for Telematics and Information Technology, 2005.