Position Paper – W3C Workshop on Semantic Web in Energy Industries; Part I: Oil & Gas

Frank Chum and Roger Cutler
Chevron Information Technology Company, USA

September 2008

Background

Chevron has been scanning the Semantic Web standards and technologies for the last three to four years. While a member of the W3C, we started by attending the Semantic Web Health Care and Life Sciences Interest Group to learn how they apply Semantic Web technologies in their domains. We were later involved in the Semantic Web Interest Group and the Semantic Web Education and Outreach Interest Group (SWEO), contributed to the Semantic Web Use Cases and Case Studies, as well as the Semantic Web FAQ collections. We have also contributed to the W3C Advisory Committee via many face-to-face and annual meetings. The associations with W3C have brought us many opportunities to learn from the best and the brightest in the Semantic Web community and to help us to build our organizational capabilities for Strategic Research in this area. We have been working on a few exploratory pilots along the way. Our intent is to share our lessons learned through this Workshop and gain insights from others in the possible approaches we can take to overcome some of our key challenges described in this paper.

Problem Domains

The most immediate problem domains we see in our industry include the following:

1) **Semantic reconciliation of enterprise metadata**
   Metadata is scattered throughout the enterprise in unstructured documents, semi-structured spreadsheets, databases, file folders, etc. They are created by applications such as those used in geosciences or as by-product of processes such as drilling, and may be stored in databases or file systems managed by groups or individuals. Semantic reconciliation provides an organized approach to metadata management by resolving differences in meaning in order to enhance metadata shareability and interoperability.

2) **Standardization for information exchange between enterprise and business partners**
   Major Capital Projects in the Oil and Gas industry involve the sharing or exchanging of equipment specifications and information such as request for quotes, purchase orders, vendor drawings and models between enterprise and business partners. In order to reduce rework and enhance efficiency in the design, fabrication, delivery and installation of these assets , standardizations are needed for semantic reconciliation of definition and specification of each component.

3) **Information integration and delivery**
   Today's enterprises have evolved into silos of disparate applications and databases resulting in the diversity of identifiers and definition for the same entities across different data sources. Connecting information from highly diverse sources and having a shared, common understanding of the data to facilitate enterprise application integration is a central use case for application interoperability.

An Exploratory Pilot

We built an exploratory pilot using semantic technologies to link technical data and document data within a small demonstration subset of our Unix file systems in order to examine the kinds of questions we might be able to answer using these techniques. We created scripts to parse the Unix directories and file paths to extract and infer metadata, and using TopBraid Composer we developed ontology models for these file directories including such concepts as SeisWorks projects, users/groups, country/region. We loaded these triples into a Siderean Seamark RDF store coupled with a faceted search facility. We were able to merge the ontology models and perform automated inferencing to answer questions like:

- What data do I have for (project, geographical reference, asset, well…)?
- Who worked on what and when?
- … etc.

Using this approach allowed us to explore a small subset of the Unix file system's metadata in a flexible, extensible and adaptive manner. More practical applications along these lines might involve exploring a much larger part of the Unix file systems using a more robust and scalable RDF repository and combining with other metadata through different ontology models in order to provide more sophisticated opportunities for automated inferencing.

Key Challenges

There are still a number of key challenges for us to overcome. We would like to mention a few of them below:

- Domain ontologies can be complex and require major commitments from industry domain experts to build by hand. Automated ontology construction approaches could be helpful. Ontology management can also be complex as the knowledge base continues to grow.
- How to use open ontologies? Where does intellectual property start and end?
- How to promote an information sharing mindset and rationalize to a common, shared ontology? Determining the "semantic distance" between terms segregated across different applications and databases to achieve consensus in a common ontology.
- Context based mining and automatic extraction of metadata from structured and unstructured data.

The Importance of Vendor Products

We think it is very unlikely that hands-on expertise with the details of Semantic Web technology will become widespread in companies such as ours. This means that we are likely to be highly dependent on vendor products to provide solutions based on this technology. It is very encouraging to see a rapid growth in Semantic Web vendor offerings, many of which have started in the Health Care and Life Sciences industry and are now looking for new markets.

Conclusion

Delivering machine-understandable semantic information across the enterprise can provide opportunities for enhanced application interoperability, automated information discovery, and execution of more contextually relevant searches. We need to be vigilant in asking the following key questions when considering Semantic Web standards and technologies:

- What does the Semantic Web bring to the table that cannot be solved by proven conventional technologies? How does automated inference help solve the business problems?
- Where does the needed metadata come from? Upfront effort is usually needed to classify, categorize, tag and extract meaningful metadata for semantic processing.

Based on the examples of the W3C Health Care and Life Science Interest Group and other W3C Semantic Web Working Groups, we believe establishing a W3C Interest Group on Semantic Web in the Oil and Gas Industry could provide substantial benefit to the industry by providing a framework for discussion forums, shared learning, exploration, collaboration and implementation in the context of the world-class expertise in the field resident at the W3C.