Towards a User-Focused Approach to the Semantic Web

A Position Paper for the International Semantic Web Workshop

Scott Tsao and Christina Portillo
Boeing Commercial Airplanes
July 30-31, 2001

Introduction

The announcement of the World Wide Web Consortium (W3C) Semantic Web (SW) Activity has drawn tremendous interest around the world since the beginning of this year. Most of the discussions on this topic tend to be "technology focused," i.e., interest or working groups formed among technicians familiar with a certain set of standards or technologies with the intent to explore and justify their role in the (vaguely-defined) "Semantic Web." Ironically, coming from the user community of the (yet-to-be-developed) SW technologies, we often find it difficult to be engaged in any of those groups because we lack prerequisite knowledge of the local "dialect." For the SW initiative to be successful, The Boeing Company believes that a "user-focused" approach is required. As such, we would like to share our view of the SW primarily from a user's perspective.

Since the fall of the "Tower of Babel" we as human beings have been obsessed with the desire to communicate with each other. Now that the World Wide Web (WWW) is a reality in our daily lives, the same obsession motivates us to find better ways to communicate via machines that are Web-aware. For the purpose of this paper we categorize our potential use of the Semantic Web in the following four (not necessarily mutually exclusive) areas:

- Data Interchange
- Application Integration
- Information Sharing
- Knowledge Discovery

Data Interchange

XML-based data interchange normally requires the use of an XML-based information model definition mechanism (i.e., DTD or Schemas). However, as the number of DTD's and Schemas increases, along with the variety of underlying information model definition languages (e.g., W3C XML DTD, W3C XML Schema, and RELAX), we see an emerging need for standards and technologies enabling "semantic convergence" among those information model definition mechanisms. One approach is to segregate the DTD's and Schemas into domain-specific repositories that can be looked up through federated registries, and to define common structures, services, and behaviors. For example, the CommerceNet eCo Framework Project defines a seven-layer architecture of collaborating registries from "Networks" all the way down to "Information Items." We believe that SW-enabled registry services facilitating automated index and intelligent query through controlled vocabularies are needed to bring the WWW from "semantic dispersion" into "semantic convergence." And we expect that coordination between the W3C SW

Activity and XML Protocol Activity will leverage and harmonize similar initiatives such as UDDI and ebXML.

Application Integration

XML-based application integration normally employs a mapping mechanism between an application's native meta-model definition languages (e.g., UML, EXPRESS, SQL) and an XML-based meta-model definition language (i.e., DTD or Schemas). One approach is to specify the mapping rules between two different languages (e.g., UML and DTD), as exemplified by the OMG XMI specification for interoperability between CASE tools. The ISO 10303 STEP Part 28 Project (XML Representation of EXPRESS Schemas and Data) is taking a similar approach to enable the interoperability between CAD/CAM/CAE tools for product design and engineering and other applications used within the product life cycle. For example, the PLCS Initiative "seeks to provide global agreement on the definition and communication of the information needed by users to plan and execute support for complex, long life assets such as aircraft, ships, and large industrial plants." We believe that SW-enabled mapping services facilitating the discovery, merger, and transformation of meta-models underlying various applications used in the product life cycle are needed to accomplish the PLCS vision.

Information Sharing

While Data Interchange and Application Integration are required primarily for structured information, a significant portion of valuable information assets is natively stored in semi-structured or unstructured format. Any addressable piece of data on the Web could be a piece of the puzzle needed to complete the whole picture of the sharable information asset. Ever since publication has been available in the civilized world, human beings have been employing various ontological representations (e.g., taxonomy, thesaurus, index, and context rules) to facilitate the process of "piecing together the puzzle." The basic approach we are taking even in the Internet age today is not much different. For example, the portal interface to the Web allows people within a given community to organize information in terms of customized portlets that can be navigated using a pre-defined taxonomy, and more advanced search and navigation can be accomplished through a built-in thesaurus or index. When information sharing is required across multiple communities, semantic mapping is required between multiple ontological representations. We believe that SW-related standards and technologies that enable the construction, merger, and exchange of such ontological representations, independent of the Web information resources, are needed to facilitate advanced search and navigation across community boundaries. In addition to XML-based linking technologies such as XLink (along with linking features available in DTD, RDF and XML Schema), we expect the W3C SW activity will leverage accomplished works such as ISO 13250 Topic Maps to accelerate the availability of standards and technologies for Information Sharing.

Knowledge Discovery

The ultimate goal of the practice of knowledge management is to make tacit knowledge of human beings explicit and reusable. One approach is to perform text mining and data mining of any "raw" (in whatever shape and form) information available using various (often sophisticated) algorithms, and to perform "sense making" using various visualization, clustering, and affinity analysis techniques. This is an interesting research area that has not seen significant employment of XML-based standards and technologies. We encourage active cross-pollination between those related communities to explore, identify, and develop SW technologies to facilitate Knowledge Discovery.

Glossary of Acronyms

CAD	**C**omputer-**A**ided **D**esign
CAE	**C**omputer-**A**ided **E**ngineering
CAM	**C**omputer-**A**ided **M**anufacturing
CASE	**C**omputer-**A**ided **S**oftware **E**ngineering
DTD	**D**ocument **T**ype **D**efinition
ebXML	**e**lectronic **b**usiness **XML**
ISO	**I**nternational **S**tandards **O**rganization
OMG	**O**bject **M**anagement **G**roup
PLCS	**P**roduct **L**ife **C**ycle **S**upport
RDF	**R**esource **D**escription **F**ramework
RELAX	**RE**gular **LA**nguage description for **X**ML
SQL	**S**tructured **Q**uery **L**anguage
STEP	**ST**andard for the **E**xchange of **P**roduct Model Data
SW	**S**emantic **W**eb
UDDI	**U**niversal **D**escription, **D**iscovery, **I**ntegration
UML	**U**nified **M**odeling **L**anguage
W3C	**W**orld **W**ide **W**eb **C**onsortium
WWW	**W**orld **W**ide **W**eb
XLink	**X**ML **Link**ing Language
XMI	**X**ML **M**etadata **I**nterchange
XML	e**X**tensible **M**arkup **L**anguage