THE ROLE OF SEMANTICS AND INFERENCE IN THE SEMANTIC WEB *A COMMERCIAL CHALLENGE*

Rob Jasper

Anita Tyler

{Rob, Anita}@HawthornTech.Com
Hawthorn Technology Group, LLC
524 Second Avenue, Suite 500
Seattle, Washington 98104-2323

Rob Jasper is VP of Research and Technology Assessment and Anita Tyler is VP of Technology at Hawthorn Technology. They are currently working to provide a foundational architecture to a new company focused on enterprise knowledge sharing and awareness using the semantic web. Both are former members of the technical staff at FizzyLab and former researchers at Boeing's Research and Technology center in Seattle.

Many of us believe in Tim's, Jim's and Ora's semantic web vision in which "a new form of web content that is meaningful to computers will unleash a revolution of new possibilities." This vision relies on transforming web content from a form that facilitates human interpretation (i.e., HTML) to a form that's designed for interpretation, reasoning, and manipulation by computer agents (based on XML). We believe agents will eventually become the prime consumers of raw web content, acting as knowledge intermediaries, transformers, and brokers, for the end consumers, humans. In this vision, humans are consumers at the top of a "knowledge food chain" that rests upon a foundational layer of semantics and inference technologies. Despite all the rhetoric about the importance of semantics, and to a lesser extent, inference, it's not entirely clear what role semantics and inference play in today's web or tomorrow's semantic web. Different researchers see different roles for semantics and inference in the semantic web. Some questions to consider are:

- How do "intelligent" programs running on the web today function without an underlying layer of semantics and inference?
- To what extent does XML itself convey semantics?
- Does semantics play any role in the runtime operation of the semantic web or does it function primarily as a software engineering tool?
- Can the semantic web exist without inference?
- What needs to be done for these technologies to prosper commercially?

We are currently working to provide a foundational architecture to a new company focused on enterprise knowledge sharing and awareness using the semantic web. After spending years in a research setting we've taken a fresh look at the world from a more commercially-oriented perspective.

Semantics Today

An increasing percentage of web pages are dynamically generated from structured or semi-structured information sources (*e.g.*, databases, knowledge bases). Because the rendering of these pages occurs in HTML, the underlying structure and linkage to meta-data is lost. Many commercial applications attempt to recover the underlying "semantics" through screen scraping and technologies like wrapper induction. With all the problems and brittleness of this approach, it continues, demonstrating commercial demand for access to the structure and semantics of the underlying data.

Standards such as XML provide a convenient mechanism for bypassing the HTML layer and connecting producers more directly to consumers of semantic content. Nevertheless, XML by itself is not capable of conveying semantics. So why should using XML make the job of any easier? XML only conveys semantics to the extent that humans representing producers and consumers agree ahead of time on the meaning of the tags. Programmers, then embed their agreements into the programs that act as producers and consumers of XML content.

Standards such as XML Schema function as software engineering design specifications for programs and play a role in checking whether XML data is well formed or not. XML Schema does not convey the semantics of the data; the semantics must be agreed to at design time.

Semantics Tomorrow

Languages such as DAML+OIL and to a lesser extent, RDF Schema provide the means of describing semantics in a way that is machine interpretable. The fact that these languages are machine interpretable and have semantics does not imply that they will have any role in the runtime operation of the semantic web. These languages and their associated inferencing procedures could be used simply as a means to the same end as XML Schema. That is, they will play the role of sophisticated software engineering tools to help humans come to agreements about semantic exchange. Some researchers have argued that this might be the only role for semantics and inference in the semantic web. We disagree. While we believe this role is important, restricting semantics and inference to a software engineering role significantly limits the potential of the semantic web. It continues to leave developers with the problem of having to obtain *a priori* agreement between parties wishing to have a "semantic exchange". While the semantic web can exist without runtime inference, the benefits of semantics and inference as purely software engineering tools aren't enough to warrant significant investment by commercial enterprises.

The Challenge

We believe that runtime application of semantics and inference is an important factor in commercial adoption of semantic web technologies. An example of a technology desired by the commercial sector is automated generation of "glue functions" that can translate terms between related ontologies. Automated translation will free developers from having to embed hard-coded knowledge of the semantics of every meaningful tag into their applications. Applications will be able to reason about new terms provided they can relate them to something they already understand.

Having semantics and inference play a significant runtime role in the semantic web creates a different set of challenges than simply relying on them as software engineering tools. Runtime interpretation of semantics and inference introduce new problems that aren't currently a major focus of the research community. These include such issues as:

- Performance and scalability
- Reliability
- Replication
- Incremental update of assertions and rules
- Security

We encourage the research community to join us in addressing these issues, thereby helping to bridge the gap between academia and widespread commercial adoption of these technologies.