Statement of Interest: Prof James Hendler

Director, Semantic Web and Agent Technologies
Maryland Information and Network Dynamics Laboratory
University of Maryland, USA
http://www.cs.umd.edu/~hendler

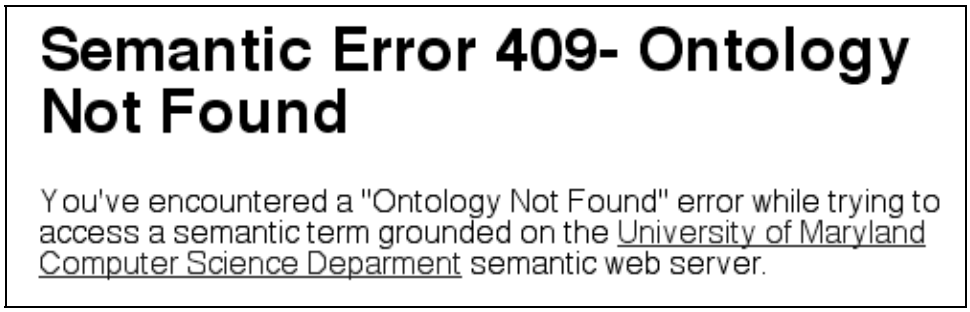
To understand this short article, I need you to use the following visualization technique. Spread your hands about one foot wide. Imagine your right hand is the traditional AI knowledge representation view of formal ontologies. Think of moving left as moving from carefully constructed large ontologies to smaller component ontologies, created in less formal ways. About the time you hit your left hand you're at the edge of the traditional AI view of ontologies, and into something much less formal and somewhat less well-defined. Now turn your head left and squint into the distance -- that's where my view of ontologies can be found!! In this little piece let me see if I can explain this, and help you realize that ontologies are going to change the world, but only if we change our view of ontologies!

It is clear to many of us [1] that the field of knowledge representation is about to get shaken up and bent out of recognition in the way that the fields concerned with on-line documents were changed a few years back. The text markup and hypertext communities had spent years working on expressive text markup languages and complicated programs designed to allow detailed analysis of online documents coupled together by two-way links between the pages. Great care was taken to guarantee that one would never hit a link that "didn't work." Guaranteeing consistency in the links required centralized repositories and complex protocols to make sure if a document was moved, the links moved with it. The community knew that the death of their systems would arise if links could be broken, and they took great pains to avoid that!

However, this assumption was blown out of the water by a small program called "World Wide Web" that was supported by a simple protocol called the "HyperText Transfer Protocol" (the "http" you have become so used to). The creator of this program, Tim Berners-Lee, realized that the hypertext systems envisioned by his peers could never scale -- if person A needed permission, or worse, had to pay, person B to link to a text, the system wouldn't grow. Instead, he developed a scheme that challenged the fundamental assumptions of the hypertext community, and his success has been legion! Further, the research communities doing markup and hypertext were profoundly changed, and current meetings have few of the "old-timers" involved as major research participants.

What is the tidal wave threatening to swamp our community? The Semantic Web effort, led by a combination of researchers from government, industry and academia, wants to take ontologies (and other KR technologies) and bring them to the web. However, this means at a web scale! When these things start to proliferate, some of our basic notions of KR will be challenged, and our community will lose if we try to kick in our heels and stay firm. For example, on the web not only is there no way to ensure consistency, but it is guaranteed that there will be inconsistent information, information one doesn't believe, information that one cannot find the provenance of, information you put there being combined with information that others described (and which you may not approve of) and all those other "web" things. Even worse, if I point at terms in your ontology, and then you change it (or move it somewhere else) my representation becomes ungrounded -- KR has never before had to deal with the AI equivalent of Error 404!

What will this brave new world of



ontological information look like? Consider the current web and the way it works. Documents of many sizes and shapes are available all over the place, with some sites developed just to index others (some of the most popular sites on the web). These documents are linked in many different ways, and they refer to things on other pages with little or no effort required. Web ontologies will grow in that same way. People who know that ontologies exist (the "web masters" of the semantic web) will help define ontologies for areas of interest to them. These ontologies will use terms from other ontologies and so on. In addition, products written for end-users, powered by access to these ontologies, will also allow for the extension of ontological terms (for example, a form to be filled out might have a field called "other" which, when named, asserts a new property of the class being extended).

Now, with this in mind, let's look at ontology languages. First and foremost, a web ontology language must be anchored on the web. This means that each "symbol" must be grounded at a "universal resource indicator" (URI). Thus, on the web there is no such thing as a generic "person" filed, rather one must refer to a particular definition such as http://www.cs.umd.edu/projects/plus/person-ontology.daml #person" which is in turn anchored to other web locations. This notion of web embedding is absolutely crucial to the semantic web, and thus it is a critical property.

On top of the is the ability to relate these grounded terms one to another. Thus one can create classes and subclasses, use descriptions or subsumption, etc. However, it is crucial to keep in mind that this is the web, and that the web is huge! No reasoner can work by assuming it knows everything, has found all possible classes, etc. without being able to assert what it is closed with respect to (a particular document or site, the collection of facts found by a semantic web crawler, etc.) Further, it is worth remembering that this is the web – different people will use the same semantic assertions for different uses, and thus the exact form of a web language may not matter as much as the fact that everyone uses it!

I'll even take that a step further – the semantic web will not be powered solely by "reasoning systems," but by a wide variety of tools and techniques that use ontologies as their interlingua. These very different agents will need to be able to trace terms found on different pages and find common referents – the content will be distributed, inconsistent, and preferably (ready for this?) not very expressive!! On the web expressivity is the kiss of death – simpler languages and solutions go further than complex ones – HTML, a watered down version of the much more expressive SGML, took over the world. SGML developers are still wandering around telling people how everyone else is doing it wrong, but strangely no one seems to listen.

And that, I believe, is the lesson we in the KR community must take to heart. Web-based ontologies are poised to move out to the web and become embedded in web tools (that is, most users will not even know they exist as they take advantage of the ontological content included in shrink-wrapped programts ranging from specialized process modeling tools like UMLS to home page creation tools brought shrink-wrapped over the counter and business-related applications like spread sheets and word processors. Note also that AI is already present in many of these tools, but the users aren't aware of it as they mouse the green lines to fix grammatical errors or use rule-based tax preparation systems.) When this happens, KR researchers will have a choice – move with the times, or become the old ignored grouches complaining about how the world simply doesn't understand what this KR stuff is all about.

Tim Berners-Lee, Jim Hendler and Ora Lassila, The Semantic Web, Scientific American, May, 2001.