

# Research Statement

Jeff Heflin

My current research interests lie in the Semantic Web, a growing research area where artificial intelligence is applied to the Internet with the goal of providing agents with the ability to understand web pages. Researchers in the field of knowledge representation (KR) have studied techniques for storing, modifying, and reasoning with complex information for decades; however, certain properties of the Internet, such as its size, distributed nature and rapid evolution, make it a challenging new domain for such work. A particularly relevant subfield of KR is the study of reusable knowledge components called ontologies. Recent research has shown that semantically marking up web pages using terms from an explicit ontology can greatly improve retrieval, integrate the data of many pages, and enable intelligent internet-based agents as well.

There are many issues to consider when designing an ontology language for the unique environment of the Internet. First is the inferential capability of the language. Should the language support taxonomic reasoning, first order logic, non-monotonic reasoning, probabilistic logic, or something else? How can the reasoning methods be scaled to the immense size of the Web? Since there is no single controlling authority, contradictions will be inevitable in most languages. How will they be resolved? How will we enable semantic interoperability, that is, the ability to integrate information that is owned by different parties who have distinct vocabularies and conventions? How can the ontologies evolve over time without breaking any dependencies on them?

One must also consider how such an ontology language could actually be used on the Internet. There are social issues involved with encouraging users to describe their documents in the language. What tools and techniques can be used to make this a less onerous process? There could be an enormous number of ontologies. How can users find whether an existing ontology meets their needs? How will trust and authority be accounted for? How can users quickly and easily formulate queries that span ontologies?

In my work, I have explored potential answers to many of these questions. I am one of the co-designers of an internet ontology language called SHOE and have developed a prototype system in order to demonstrate its feasibility. This system includes a tool for assisting users in the annotation process, a web-crawler that gathers the markup from web pages and stores it in a knowledge base, and a number of query interface tools. Demonstrations of these tools are available online at <http://www.cs.umd.edu/projects/plus/SHOE/>.

There is evidence that continued work in this area is likely to generate funding. SHOE has been used in projects for the FDA and Intelink, the U.S. Intelligence community's intranet. Even more importantly, it is one of the chief inspirations for the \$50 million DARPA Agent Markup Language (DAML) effort, to which I volunteer time by serving as a member of the Language Committee.

Future research directions include applying the distributed ontology approach to interagent communication. Although the need to handle different types of speech acts increases the complexity of the problem, it gives agents the ability to ask questions in order to resolve ambiguities or verify correct understanding. Other areas of future interest include the integration of information extraction, natural language processing, and machine learning into the knowledge acquisition process, in order to reduce the amount of human effort needed to mark up documents.