Towards a Semantics for the Web

Christopher Welty[†]

Vassar College Computer Science Dept.
Poughkeepsie, NY 12604-0462, USA
on sabbatical at LADSEB/CNR
Padova, Italy
weltyc@cs.vassar.edu
http://www.cs.vassar.edu/faculty/welty/

Abstract

The popularity and press surrounding the release of XML has created widespread interest in standards within particular communities that focus on representing content. The dream is that these standards will enable consumers and B2B systems to more accurately search information on the Web within these communities. We believe the expansiveness and diversity of the Web creates a need for a small set of standard semantic primitives that have the same meaning and interpretation across communities. Such a standard set of primitives should take into account existing efforts in ontology, and in e-commerce content standards. We are investigating existing content standards proposals for the Web, and present some basic motivations and very preliminary ideas regarding what such a standard set of semantic primitives could be. I begin with some quotes from the workshop so far, and then present some examples of work in the library and e-commerce domains, and how they might be harmonized.

Related Comments

Some comments from other talks at this workshop:

- *The problem is not that there are no semantics, the problem is that the semantics is hidden in software components* - Stefan Decker.
  Additionally, note that hiding semantics within software components creates interoperability problems. When the semantics is hidden, it is far more difficult to find, to understand, and to integrate. Our goal is to make this semantics explicit, in a declarative way, with as little ambiguity as possible. See (Guarino, 1998b).
- *Form and content are different things and should not be represented with the same tags* - Wolfgang Wahlster.
- *Ontology mapping is very important, but we don't know how to do it* - Guus Schreiber.
  There is a difference between "knowing how" to do something and actually doing it. The latter requires having the proper underlying technology support, and people to use it. I claim we do know how to do ontology mapping, but at the moment it is an intensive task that requires a lot of human effort. See (Guarino and Welty, 2000) for the latest on this.
- *Communities will form their own ontologies* - Jim Hendler.
  And clearly they are, and have. Libraries, for example, have ontologies that date back several hundred years. Lawyers have "best practices" about a hundred years old that are captured and represented, albeit in text. We are fooling ourselves if we believe we're doing anything new. We can learn a lot from other communities. See (Guarino, 1998a) and (Welty and Jenkins, 1999) for more on this.
- *Are there language requirements for mapping ontologies* - Nicola Guarino
  This seems to have been the focus of the workshop, as opposed to semantics for the web.
- *KR language = formalism + ontology* - Richard Fikes (through Nicola Guarino)
- *Lightweight ontologies good* - Mike Uschold
  An emphasis on small ontologies that can be easily agreed upon and deployed quickly has paid off.
- *Scale...* - Jim Hendler
  The web is big. If we really want anyone to pay attention, we better be able to deal with its scale. This makes the last comment by Mike even more pertinent.

Ontology

The good news, for me, is that most of the workshop participants seem to have taken for granted that ontology is necessary to have web semantics, however much of this workshop has focused on the *form* rather than the *content* of these ontologies. I will focus on the form, and so I require making a bit more explicit what an ontology is.

Nicola Guarino (1998b) defines an ontology to be an implemented artifact that attempts to constrain the intended meaning of a vocabulary by eliminated *unintended model*s in the interpretation. In (Welty, et al, 1999) we defined a spectrum of ontology kinds, differentiating them only in terms of their *ontological depth*, that is, in Guarino's terms, the degree to which they eliminate unintended models. These are, in order of depth:

1. Lexicon (Vocabulary with NL definitions)
2. Simple Taxonomy
3. Thesaurus (Taxonomy plus related-terms)
4. Relational Model (Unconstrained use of arbitrary relations)
5. Taxonomy and relational model. (Type restrictions and is-a links, some notion of inheritance)
6. Fully Axiomatized Theory

If we accept all these as types of ontologies, then we are forced to admit that it is not how ontologies are represented that is significant, but what they represent.

Existing Ontologies

Other communities have ontologies, perhaps they don't call them that, but they fit into the definition given above. We can learn a lot from these other communities – but that's not to say they can't learn from us, too.

Libraries have had three interesting ontologies for a long time (Welty and Jenkins, 1999), though all of these are quite fundamentally affected by digitization and the web:

- The card-catalog ontology, which has come to define meta-data (author, title, publisher). This ontology is roughly five hundred years old (at least).
- The bibliographic ontology, which defines records for articles inside periodicals and other documents. This ontology is about 100 years old.
- The subject ontology, which carves the world into discrete subject areas. This ontology is about 150 years old, though has been under constant evolution during that time.

Other ontology efforts have been ongoing for some time, or have just started as a result of the hype surrounding XML and the desire to support e-commerce. For example:

1. *General* content standards and ontologies (WordNet, CYC, ISO/BSR, CALS/UDEF, …)
2. *Process* standards (NIST/PSL ontology, DARPA/CPR, …)
3. *Product* standards (ISO/STEP, UN/SPSC, RosettaNet…)
4. *Information* media standards (Dublin Core, INDECS, CIDOC…)
5. Conceptual modeling and *representation* standards (UML meta-model, EPISTLE…)

See (Guarino, Welty, and Partridge, 2000) for a larger sampling, though still incomplete.

Learning from other efforts

A survey of these other efforts can help to answer questions about what is relevant to creating a semantics for the web. There are many threads than run through a lot of these efforts and questions that a lot of communities seem to be asking themselves in isolation:

- What is a person
- What is a document
- Document components
- Paper vs. electronic documents
- Events and states

There is widespread confusion regarding the use of subsumption and taxonomies for organizing information. See (Guarino & Welty, 2000) for more on this.

There is widespread confusion on how to use the part-of relation, see (Artale, et all, 1996).

Benefits

Investigating these areas can have important consequences:

1. It can solve a host of existing problems, making it possible for new efforts to get results faster. Most new efforts tend to get bogged down in the same terminological problems at the start.
2. It can enable interoperation between communities. Most people seem to believe by having a common form, (i.e. XML) they will be able to interoperate with anyone.

Towards a Standard Semantics

I take for granted that there should be a *standard* semantics for the web. This may seem controversial for two reasons:

1. Standards are difficult to achieve and to be accepted
2. Standard semantics sounds like top level ontology.

I specifically avoid using the term “top level” ontology here because it is associated with a large, monolithic, brittle, view of the world.

On the contrary, a standard web semantics should be small & lightweight, domain independent, and should be inclusive rather than exclusive. It should evolve from existing work in many different fields, not developed in isolation and thrust upon the world.

References

Artale, A., Franconi, E., Guarino, N., and Pazzi, L. 1996. Part-Whole Relations in Object-Centered Systems: an Overview. *Data and Knowledge Engineering*, **20**(3): 347-383.

Guarino, N. (ed.) 1998a. *Formal Ontology in Information Systems.* Proceedings of the First Conference (FOIS'98). Trento, Italy, 6-8 June 1998. IOS Press.

Guarino, N. 1998b. Formal Ontology in Information Systems. In N. Guarino (ed.) *Formal Ontology in Information Systems.* Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. IOS Press, Amsterdam: 3-15.

Guarino, N., and Welty, C. 2000. *A Formal Ontology of Properties*. LADSEB/CNR Technical Report 01/2000. Available from http://www.ladseb.pd.cnr.it/infor/ontology/Papers/OntologyPapers.html.

Guarino, N., Welty, C, and Partridge, C. 2000 Towards Ontology-based harmonization of Web content standards. *Proceedings of the ER-2000 Workshop on Conceptual Modeling Approaches for e-business.* Available from http://www.ladseb.pd.cnr.it/infor/ontology/Papers/OntologyPapers.html

Welty, C. and Jessica Jenkins. 1999. An Ontology for Subject. *J. Knowledge and Data Engineering.* **31**(2)155-182. Elsevier.

Welty, C., McGuinness, D., Uschold, M., Gruninger, M., and Lehmann, F. 1999. *Ontologies: Expert Systems all over again?* AAAI-1999 Invited Panel Presentation. Available at http://www.cs.vassar.edu/faculty/welty/aaai-99/.