An Overview of Machine Ontologies for Ethologists

Peter E. Midford
April 2002

Introduction

This is an overview based on about a year's experience with building ontologies for behavior and trying a small number of tools. It certainly reflects the bias of my experiences with particular tools (e.g., Protégé and Ontolingua) and projects (e.g., GeneOntology). I hope this will encourage people to look at ontologies for behavior from a variety of approaches and perspectives.

What is an ontology?

Originally "ontology" referred to a branch a philosophy focused on questions of existence. It was borrowed in the 1980's by the artificial intelligence community to refer to systems for representing and using commonsense knowledge about a domain. Grubb (1992) provided a commonly cited, if rather opaque definition, "a specification of a conceptualization."

How would ontologies be useful in the study of behavior?

1. Ontologies are successfully in use to link terms across independent databases. For example, the GeneOntology currently links terms across 12 taxon-specific gene databases and 6 gene product databases.
2. Most or all the information in ethograms can be represented using an ontology and possibly an associated knowledge base of behavioral events.
3. Ontologies could allow new methods of analysis (I am working on comparative methods for ontologies) and "smart", flexible data-collection tools. If we ever get intelligence observation systems (direct from video to behavior) some type of ontology will likely be present.
4. If you believe the hype, ontologies are the future of web-publishing (e.g., Berners-Lee et al. 2001, McGuinness 2002).

What types of Ontology are there?

In practice, the word "ontology" has been applied to systems along a wide range of complexity, including:

- *Controlled vocabularies*

  These are simply lists of words or phrases, representing a vocabulary that users agree upon. There are many examples from medicine. A checklist of behavior patterns would also be an example.

- *Hierarchical classifications of terms*

Structure based on taxonomy ("isa") or based on several relations (isa, part-of, etc.) The structure consists of a tree or graph (network) with words at the nodes and relations represented as links between words.

- *"Frame" – based systems*
  A frame is a node with a fixed set of named relations (called slots). The set of slots is usually established by some sort of inheritance mechanism. Frames are similar to classes in object-oriented programming.
  - Example: The frame below contains two slots specific to its type/class (isa and sensu) and several slots that would be filled in by *instances*, which represent particular symmetric_pair_wave events

```
Symmetric_Pair_Wave
   isa: Symmetic_action
   sensu: Habronattus californicus
   start_time:
   end_time:
   instrument:
```

- *Logic* – The most expressive and powerful, but generally not the easiest to use. Descriptive Logics are the most popular variety of logic for ontology construction.
  - Example

```
Third_pair_wave(event1)
Second_right_leg_step(event2)
Start_time(event1,time20)
Start_time(event2,time21)
End_time(event2,time22)
End_time(event1,time23)

and(Start_time(?e1,?t1))
     (Start_time(?e2,?t2))
     (Before(?t1,?t2))
=>
Starts_before(?e1,?e2)
```

What are some examples of Ontologies?

1. "General" Ontologies

Besides the ontologies listed below, there are several web pages listing ontology projects (e.g., http://www.cs.utexas.edu/users/mfkb/related.html).

Cyc (http://www.cyc.com) One of the largest and best known knowledge bases, containing over 1 million logical assertions. A chunk of Cyc and its Cycl language are in the process of being released to open source.

WordNet (http://www.cogsci.princeton.edu/~wn/) A lexical database of over 138,000 English words linked into a network.

2. Biological Ontologies

Numerous ontologies have been developed for domains in medicine and molecular biology.

EcoCyc/MetaCyc (http://ecocyc.org/ecocyc, Karp et al. 2000) One of the first biology focused ontologies, EcoCyc is a network representation of metabolism in *E. coli*. MetaCyc represents metabolism not specific to *E. coli*.

GeneOntology (http://www.geneontology.org) A network of over 10,000 terms related to gene products. Organized as three hierarchies of terms: molecular function, biological process, and cellular component.

Microarray Gene Expression Data group (http://www.mged.org) This project includes an ontology related to microarray data. The ontology is rapidly developing as of March 2002.

3. Ontologies for Behavior:

Cyc has released an "upper level" ontology (at http://www.cyc.com/cyc-2-1/index.html), which contains several terms for animal behavior.

GeneOntology has a small section of its Biological Process tree devoted to behavior.

My own ontology project (http://www.mesquiteproject.org/ontology/) includes an ontology coded from a published ethogram of sea turtle nesting (Hailman and Elowson 1992) and some material related to jumping spider courtship.

Other Approaches to Ontology Building

Although term networks,frames and description logics are the dominant approaches to ontology construction, there are others. In particular "conceptual graphs," developed by Sowa (e.g., Sowa 2000) may be an attractive combination of visual representation with expressiveness .

Tools for ontology building

In addition to the list below, Duineveld et al. (2000) provides a somewhat dated overview of several tools.

Protégé-2000 (http://protege.stanford.edu) A very popular and user friendly ontology editor and knowledge-base entry tool. The web site also has numerous example ontologies and plug-ins, along with instructions for getting started (Noy & McGuinness 2001).

Ontolingua(http://www.ksl.stanford.edu/software/ontolingua/, Farquahr et al. 1997) A web-based tool for ontology construction that supports collaboration. Offers a large library of ontologies as examples and for reuse.

Dagedit (http://sourceforge.net/projects/geneontology/) This is an editor created by the Berkley Drosphilia genome project for the GeneOntology consortium. It edits DAGs (directed acyclic graphs) the network structure used by the GeneOntology.

DAML+OIL (http://**www.daml.org**) This is a language, not an editor, but it may represent a future standard for ontologies. This XML-based language is the combination of two previous projects. Currently the only freely-available editor is OilEd(**http://oiled.man.ac.uk/**), which is limited to small ontologies (the notepad of ontology editors). Other systems are promised soon.

References

Berners-Lee, T., Hendler, J. & Lassila, O. 2001. The Semantic Web. Scientific American 284(5):35-43.

Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B. and Benjamins, V.R. 2000. WonderTools? A comparative study of ontological engineering tools. *International Journal of Human-Computer Studies* 52:1111-1133.

Farquhar, A. Fikes, R. & Rice, J. 1997. The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human Computer Studies*. 46:707-727.

Gene Ontology Consortium 2000. Gene ontology: Tool for the unification of biology. *Nature Genetics* 25:25-29.

Hailman, J. P. & Elowson M. 1992. An ethogram of the nesting loggerhead sea turtle. *Herpetologica* 48:1-30.

Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. & Pellegrini-Toole, A. 2000. EcoCyc: Electronic Encylopedia of *E. Coli* Genes and Metabolism. *Nucleic Acids Research* 28:56.

McGuinness, D. 2002. Ontologies come of age. In: *Spinning the Semantic Web* (Ed. by D. Fensel, J. Hendler, H. Lieberman & W. Wahlster) Cambridge MA:MIT Press (in press).

Noy, N. F. & McGuinness, D. L. 2001. Ontology Development 101: A guide to creating your first ontology. Knowledge Systems Laboratory Tech report KSL-01-05 Stanford University.

Sowa, J. F. 2000. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove CA: Brooks/Cole.