BY JOHN MICHON, MD, MS, DUKE UNIVERSITY SCHOOL OF MEDICINE

Ontologies: A Way of Capturing and Representing Knowledge

Ontology, defined as "a theory of existence," has an important place in Western philosophy. It seeks to account for all phenomena in the universe.

In biomedical computation, however, an ontology is narrowly defined as "a description of the concepts and relationships that can exist in a domain." Specifically, ontologies define what *things* may comprise a domain, their *properties* and how they *relate* to each other.

Ontologies are useful for organizing and structuring knowledge so that it can be *stored*, *shared* and *reused*. More than a simple, static repository of knowledge, ontologies can be used to infer and create new knowledge. They may range from fairly simple hierarchies or taxonomies to logical theories that embody rich, complex knowledge.

As a simple example, consider an ontology of a human family. Basic concepts would include parent, child and sibling as well as more complex ones such as stepson and second cousin. Rules for such an ontology would include "each parent has at least one child" and "each child has exactly two biological parents." Certain concepts could have symmetric (brother and brother) or inverse (parent and child) relationships. Logical inference can be performed such that if we know Peter and Jennifer (male and female) share the same two biological parents, we can infer that they are brother and sister.

These same principles can be amplified in larger, more complex ontologies that cover broad biomedical domains. For example, the Foundational Model of Anatomy (FMA), which covers gross and microscopic human anatomy (from organ systems to cellular and molecular structure), contains approximately 72,000 concepts and over 113,000 descriptive terms using 168 relationship types.

As another example, the Gene Ontology (GO) dramatically improves biologists' ability to identify and associate genes involved in common cellular tasks and to compare gene function across species. GO describes gene products in terms of their associated biological processes, cellular components and molecular functions. For example the eye protein rhodopsin performs the function of *phototransduction* as part of the process of *visual perception* as a component *integral membrane protein* in *photoreceptor* cells. Representing our knowledge of rhodopsin with a series of such consistent and controlled concepts and attributes allows unified queries across databases that contain information about this and related proteins.

> A biomedical ontology must represent an area of knowledge in a way that both humans and machines can understand.

A biomedical ontology must *represent* an area of knowledge in a way that both humans and machines can understand. The recent development by the World Wide Web Consortium of OWL (Web Ontology Language) has been a major advance in this regard. OWL allows complex ontologies to be expressed based on a firm logical foundation with richly defined relationships, including inverse, transitive, complementary, symmetric and disjoint properties.

By representing knowledge in ontologies we shift the burden of routine information processing to computers, freeing humans to spend more time in creative, intuitive and imaginative activities. Moreover, as ontologies assume a growing role, genomic, proteomic and metabolomic data may be integrated with clinical information to elicit clues to the molecular basis of disease. Ontologies will therefore provide a foundation for more effective and personalized medical care. □

DETAILS

John Michon, MD, MS, is in the Department of Ophthalmology at the Duke University School of Medicine. His research interests include ontologies and knowledge representation, bioinformatics and molecular visual science.

URLs

Foundational Model of Anatomy. http://sig.biostr.washington.edu/projects/fm/

Gene Ontology. http://www.geneontology.org/

OWL. http://www.w3.org/2004/OWL/