Introduction to Biomedical Ontologies #1: What is an Ontology?

Script by Jennifer R. Smith

This video is the first in a series of tutorials about biomedical ontologies. In this video we will cover some general information about what an ontology is, while later videos will include information about the ontology annotations themselves and how ontologies can help you in your research.

Before we talk about what an ontology is, let's look at the problem that the use of ontologies is meant to solve. Consider the term "football". If a friend invites you to join in a game of football, what game are you actually going to play? If he's from the UK, he might mean "Association football" or what Americans call "soccer". On the other hand he could be referring to rugby football. If he's from the United States, he probably means "American football". This example points out two problems with languages, that is, the same word can have different meanings, and the same meaning can be expressed using different words.

One way of dealing with this problem is through the use of controlled vocabularies. In a controlled vocabulary, a particular concept is assigned a single "official" term and any other terms that mean the same or very similar things are used as synonyms for that term. So this game [*show soccer players and soccer ball*] might be officially referred to as "Association football" while the terms football and soccer could be used as synonyms.

An ontology is a particular kind of controlled vocabulary. We will define it like this: An ontology is a controlled vocabulary of well defined terms with specified relationships between them, capable of interpretation by both humans and computers. It is important that each term be defined so that there is no confusion over what concept is being conveyed.

So then, what does it mean to say that ontology terms have "relationships"? Basically, an ontology is constructed as a hierarchy with the term at the top being the most general and the ones at the bottom the most specific. For example, if we were going to construct a "Vehicle Ontology" we could put the most general term "Vehicle" at the top of the "tree". Under that we might include categories like "air vehicle", "land vehicle" and "water vehicle". Then under those, an airplane would be an air vehicle, a truck and an automobile would be land vehicles and a boat would be a water vehicle, but an amphibious vehicle would be both a land and a water vehicle. So you can see that the ontology tree goes from more general to more specific. In addition, a more general term can have multiple more specific child terms and a child term can have multiple parents.

So, talking about cars and sports is all well and good, but what about biomedical ontologies? As more and more experiments in the fields of biology and medicine become high-throughput, expressing data in ways that can be read by computers, and ways that can be shared from one experiment to another or one data source to another is becoming more and more important. In response to this need, a growing number of biomedical ontologies are being developed.

Arguably the most commonly used of these is the Gene Ontology or GO, so we will use that ontology as our example. GO is being developed as a cross-species platform for describing genes and gene products in terms of their molecular functions, their biological processes and their subcellular localization or cellular components.

This is the GO Browser at MGI (http://www.informatics.jax.org). From this page you can search for a particular term or browse through the ontologies by clicking on terms of interest. If I browse down the tree in the Cellular Component ontology, you can see that the vocabulary goes from the most general

term “Cellular Component”, through “organelle” and “membrane-bounded organelle” to the more specific term “extracellular membrane-bounded organelle” and so forth.

In the AmiGO browser at amigo.geneontology.org, you can see that the term “extracellular membrane-bounded organelle” has both “membrane-bounded organelle” and “extracellular organelle” as parents.

If you’re interested in browsing through other biomedical ontologies, check out the BioPortal tool at the National Center for Biomedical Ontologies’ website, http://www.bioontology.org. This tool allows you to explore ontologies on everything from the Adult Mouse Brain Atlas and African Traditional Medicine to Yeast Phenotypes and Zebrafish anatomy and development.

So now that you know something about ontologies themselves, where can you find out what ontology terms have been assigned to a specific gene or gene product? Gene ontology annotations can be found on gene and protein records on the websites of major databases such as Ensembl, UniProtKB and NCBI’s Entrez Gene. Annotations are also found at model organism databases such as the Rat Genome Database. In fact, at RGD genes and other data objects can be assigned terms from up to four different ontologies—the Gene Ontology, Mammalian phenotype ontology, disease ontology and pathway ontology—making it simple for both people and computers to access a wide variety of functional data about a gene or group of genes.

For more information about ontology annotations and what information you can get from them, see the next video in this series: Anatomy of an ontology annotation.

For more information:

Rat Genome Database (RGD):
http://rgd.mcw.edu

RGD's complete Introduction to Biomedical Ontologies Video Series:
http://rgd.mcw.edu/wg/home/the-introduction-to-biomedical-ontologies-video-series

The Gene Ontology Consortium (GOC):
http://www.geneontology.org/

The Gene Ontology Consortium's AmiGO browser:
http://amigo.geneontology.org/cgi-bin/amigo/go.cgi

The National Center for Biomedical Ontology (NCBO)
http://www.bioontology.org/

NCBO's BioPortal:
http://www.bioontology.org/BioPortal

Mouse Genome Informatics (MGI):
http://www.informatics.jax.org/

MGI's GO Browser:
http://www.informatics.jax.org/searches/GO_form.shtml

NCBI's Entrez Gene:
http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene

The Ensembl Genome Browser:
http://www.ensembl.org/index.html

The UniProt Universal Protein Knowledgebase (UniProtKB):
http://www.uniprot.org/uniprot/