Abstract

For this project, my objectives are:

1. To map the intersection between the Cell Cycle Ontology (CCO) resources and the Gene Ontology.
2. Adapting the scripts and applying reasoning to investigate the potential of reasoning over the Gene Ontology.
3. Creating a file with all of my proposed changes to the GO and posting it on the Sourceforge.net site.

Significance

Mouse Genome Informatics (MGI) is the international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease [1]. The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism [2]. Mouse genes are functionally annotated in the MGI resource using the GO. The Cell Cycle Ontology (CCO) is a self-contained ontology focused on the representation of cell cycle knowledge [3]. The Cell Cycle Ontology has not been integrated with either the Gene Ontology or the MGI database. That means that the data and experiments used for the CCO have not been cross-referenced with data in the GO. It is possible that the data in the CCO could match that in the GO or even in the MGD. With the cell cycle (in CCO) mapped for humans, any orthologs of the human genes to the mouse genes help researchers understand cell mutations in mice during cell division and replication. These two databases merged will allow for inference and knowledge analysis on a whole new level considering the cell cycle of the mouse genome. This project is significant because it is another step in expanding and furthering the knowledge and data of the Gene Ontology and the Mouse Genome Database.

Previous Work

My project is another step in the unification of biological terms. The work already done on my project is the ontologies that I will be working with: the Cell Cycle Ontology, the Gene Ontology, and the Mouse Genome Informatics Database. The OBO Foundry is an organization that helps support and enhances these and other ontologies. In an article in *Nature Biotechnology* about the OBO Foundry, this was said, “The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'” [4]. Ontologies, however, right now are very spread out in that the all have very similar terms and knowledge and there is little conjunction and linkage between them. For instance, GO has the term 'lung development,’ the Mammalian Phenotype has 'abnormal lung development,' and the Adult Mouse Anatomy has 'lung.' All of these terms should be linked intentionally somehow, but they aren’t. They are instead only semantically linked. Here I will attempt to use published scripts and the representation of the cell cycle to examine intersections between a variety of ontological resources.

Methods

1. I will be gathering bioinformatics data. This will require that I acquaint myself with publicly available data resources. These resources will most likely include the Jackson Laboratory Library and the scientific journals there and open-source ontologies on the web.
2. I will be using Perl, a programming, scripting language, which is used very often in bioinformatics, since it has such powerful text processing methods. One of its most powerful methods are regular expressions that I will be using extensively.
3. I will be learning and using basic database concepts to create a simple database with Microsoft Access and learn SQL (Structured Query Language) for simple data queries.
4. I will be using an application called OBO-Edit for ontology reviewing and editing.
5. I will be using other GO tools to track and save my research data.

Bibliography

3. Cell Cycle Ontology. 18 June 2009 <http://www.cellcycleontology.org/home>.

2. The Gene Ontology. 18 June 2009 <http://www.geneontology.org/index.shtml>.

1. MGI-Mouse Genome Informatics-The international database resource for the laboratory mouse. 18 June 2009 <http://www.informatics.jax.org/>.

4. "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration : Article : Nature Biotechnology." Nature Publishing Group : science journals, jobs, and information. 18 June 2009 <http://www.nature.com/nbt/journal/v25/n11/full/nbt1346.html>.