Homework 4

Assigned: 10/15/08
Due: 10/22/08, hardcopy, in class

The goals of this assignment include:
- Learning more about the organization of the Gene Ontology.
- Provide an experience typical of a real world occasion; that is, you find that you have to import and/or integrate data from some existing project and some miscellaneous documents are dropped on you. Note that the Gene Ontology is among the best-documented projects you may ever have to integrate. See: http://www.geneontology.org/GO.contents.doc.shtml
- Sometimes there is no documentation, at all.

Consider the physical E-R model for the Gene Ontology database, see,


CS395T students:

Develop a logical UML model for the Gene Ontology database. Part of this was done in class. A clean version of the class presentation is attached.

CS329e students: see the next page.
CS329e students:

Your goal is to develop a logical UML model for part of the Gene Ontology database. The part of the database that concerns the representation of the three taxonomies was done in class and is attached. In this homework you are to develop a logical model for the part of the Gene Ontology database that concerns the representation of sequences, their gene products, and the association of gene products with terms in the taxonomies.

Start with the following diagram for the three strong entities, seq, gene product and term.

| seq      | gene_product | term      |

1) For each class, look to the corresponding entity in the Gene Ontology ER diagram. Fill in your logical model with (only) the logical attributes. You do not need to include the data types. By virtue of the work done in class, term has been done for you.

Please be kind to the T.A. and produce your answer using Rational Rose. Although there are many parts to this homework, the result of this homework should be a single model.

2) Add the species entity and its attributes to your diagram. Include the appropriate association(s) and the association(s) multiplicities. Hint: The Gene Ontology web document includes a legend explaining the syntax it uses to create its diagram. Do not confuse that syntax with UML syntax.

3) Add to your diagram the associations between seq and gene_product, and, gene_product and term. Hint: This may require association classes. Do not forget, the Gene Ontology web document represents a physical model. I.e. if the Gene Ontology people had started with a logical model, what you would be looking at in their documentation would correspond to the result of Rose’s first compilation step.

4) Similar to question 2, add the evidence and association_qualifiers to your diagram.