BIOL 498/CMSI 698
SPECIAL STUDIES: BIOINFORMATICS
Spring 2006

Assignment 0209

XML-to-Java-to-Relational Database Conversion

This is another "deep end of the pool" assignment, but now of a more technical nature. For this assignment, we will attempt to build a software tool chain that allows us to import XML data into Java and then a relational database as painlessly as possible.

Not for Submission

As needed, read up on one or more of these technologies as you work on the assignment — some links are on the Web site, but generally it won't be hard to find decent resources on Google or Wikipedia: PostgreSQL, JDBC, SQL, Hibernate, Xdoclet, JAXB, XML, XML Schema (xsd), HyperJAXB, Ant.

For Submission

Develop a software tool chain that imports a UniPort XML file into corresponding Java objects and finally into a relational database. Work as a group — we leave the logistics to you as a class. The final deliverable is a CD containing this software and any necessary documentation.

Functionally, the program can be separated into two phases; the first phase is concerned with database definition, and in theory need only be invoked when the UniPort XML Schema changes:

1. Accept the UniProt XSD file (available at http://www.pir.uniprot.org/support/docs/uniprot.xsd) and convert that into a set of Java classes using JAXB and HyperJAXB.

2. Run Xdoclet on the resulting Java source files to produce the Hibernate object-relational mappings from those Java classes to a relational database.

3. Create an SQL DDL file that can be loaded into PostgreSQL to set up the database.

The second phase of the program can be viewed as the database loading/retrieval cycle: once the database has been set up, XML files downloaded from the UniProt Web site should be easy to load into the system. To wit:

1. Accept a UniProt XML file, read it in, parse it, then save it to the database using a combination of JAXB and Hibernate.

2. Accept a query on the data and return the results as a Java object structure using Hibernate to access the PostgreSQL relational database.

Extra Credit

Take the UniProt XSD file and produce a UML object diagram for that schema. Of course, you will have to learn XML Schema and UML to do this properly — resources for these are easy to find online.