

BIOL 498/CMSI 698

SPECIAL STUDIES: BIOINFORMATICS Spring 2006

Assignment 0323

Time to Clean Up and “GO” On

To summarize what has happened (or not) so far, this is where things stand:

- Our primary goal for the semester is to create a piece of software that produces a GenMAPP-readable file based on at least UniProt and GO.
- Our chosen approach is to load the primary XML data (exactly as downloaded from the Web) into a relational database, then query that data and write it into the format that GenMAPP needs, which is a Microsoft Access MDB file.
- We also observe that the XML files themselves may change format over time, and so it is useful to make it as easy as possible to accommodate these changes. Thus, we are trying to build a tool that can regenerate the database schema and related code: this is XSD-to-DB.
- Certain tasks call for reusable code: (1) configure our database connection settings, (2) load XML files into the database, and (3) query the database. This is XMLPipeDB Utilities.
- Whether or not we use the prior modules, we still must be able to load UniProt XML files into a database. This is UniProtDB.
- Finally, back full circle, the whole point of all this is to make GenMAPP files — thus, at the very end of it all, we have GenMAPP Builder.

At this point, we’ve started a manual and laid out initial versions of the software. We now need to: (1) debug/clean up existing code, and (2) start working on GO, the Gene Ontology Database.

For Submission

Based on our March 16 meeting, the following needs to be done. Task allocations are based on what was discussed at the meeting:

1. *Continuing work (primarily debugging) on XSD-to-DB: Adam*

2. *Completion of XMLPipeDB Utilities (including demo): Jeffrey, Babak, David*
3. *First commits of UniProtDB and GenMAPP Builder: Joe, Joey*
4. *Initial exploration and assessment of how to add GO to XMLPipeDB: Roberto, Scott*

Now, new stuff: first, I’ve updated the manual — restructured it, made some corrections, added some material. New version has been committed to CVS. *Please read it before the next meeting.*

Here are some further notes, based on my own review and work as of March 19. I invite discussion of these, *except for the last bullet* — everyone should follow that:

- The XMLPipeDB Utilities module in CVS does not match its documented name in the manual; personally, I like all-lower-case file structures, so I prefer *xmlpipedb-utils...* which is what was in the manual to begin with.
- I propose a flat-but-versioned structure for the *lib* directory: first, remove the nested subdirectories, and second, attach version numbers to the JARs — see the updated manual for details.
- What’s the best way to ensure that no one commits broken code? When I checked out XMLPipeDB Utilities, *ant* didn’t work, and some imports were incorrect. I fixed them, but they shouldn’t have shown up in the first place.
- Finally, time is flying; we still want the software *and* a paper by the end of the semester. To pick up the pace, from now on please post *daily updates* to the developer forum. Create a thread called “*yyyyymmdd* Update” if it isn’t already there, and post to it. Just a sentence or two will do; that way if you run into any issues, we can deal with them ASAP.

OK, that’s it for now...see you Thursday!