

Final Project Reflection

Statement of Work

As the quality assurance member of the group, I was responsible for identifying the valid IDs that needed to be exported in the customized gene database for "Shigella flexneri". I first had to download all the files that were needed for TallyEngine, import them, and make a database using PostgreSQL. Doing so would allow me to later launch an export that would use those files and database to give me a tally of how many IDs exist. I was also tasked with finding IDs that exist within the UniProt XML file that were not exported into the .gdb file (that our GenMAPP Users would later use on their end). Additionally, I aided the Coder on what to put on the customized species profile and also tested each build that he made to ensure that we actually captured the IDs we need and that we did not break our existing builds. I would record the results of the builds into our Gene Database Testing Report.

I also contributed to the creation of our group's deliverables, genome paper presentation, and final project presentation.

Individual Wiki Pages:

- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Troque_Week_11
- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Troque_Week_12
- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Troque_Week_14
- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Troque_Week_15

Group Home Page:

https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Oregon_Trail_Survivors#Week_15

Gene Database Testing Report:

https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/Gene_Database_Testing_Report_-_Oregon_Trail_Survivors

Deliverables & Presentations:

- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/images/7/7a/OTS_ShigellaFinalPresentation.pptx
- https://xmlpipedb.cs.lmu.edu/biodb/fall2015/index.php/OTS_Deliverables

Assessment of the Project

I would say our group worked extremely well together. Even though there are some of us who did not understand what was happening at times, especially in terms of presenting the genome paper, the other team members would gladly help out in trying to explain the complicated jargon in the papers. We would always meet up late in the night to work together on our respective parts and/or we would communicate via text messages what we plan to do next on our project. Our project manager, Kristin, especially was very responsible in her project manager job. She would always remind us of what we needed to get done for the week so that we would finish by the deadline (although we would still scramble since we would not have time to work on them during the day due to complicated schedules). Jake and I also did not have any problems communicating, however, the underlying problem that we were both having was that the material can sometimes be confusing. Erich was also hardworking; he would stay up late with me in order to finish what we were supposed to do. Our team really helped each other out, and so we worked as a cohesive unit.

What worked for us is our communication through text messages. What didn't work is when I, myself, would get swamped with assignments for other classes such that I wouldn't really have time to work on the testing of the builds. The XML file that I downloaded also didn't quite work out for us since it was a bit different from what we used before. For starters, the ordered locus names for each gene are represented twice, but in different formats so both of those were captured by TallyEngine. In addition, tracking down the missing IDs proved a bit difficult for me since the genes existed in some other tags in the XML file.

If I could do it all over again, I would wish that I had more time so I can perform a more thorough test, but with my classes just the way they are this semester, I don't think that would make much of a difference. However, if it were the case that my assignments do not pile up as much, then I would definitely want to practice more on the presentations and work on perfecting the testing reports.

I would say that for the most part, the database is good. The only problem I have with it is that I didn't catch that the IDs should really match the Ensembl Bacteria website until much later in the project, and so not enough time was put into trying to fix that problem. The content is definitely there, and there was really no way to add more of the IDs since I tried looking for some of them manually and they really did not exist. The problem is more towards having extraneous data that could be fixed given enough time.

In terms of organization, our group was well-organized, especially since our project manager is actually doing her job of reminding us of what needs to get done. Our home page could use some minor retouches, but other than that, it is still fairly organized.

Having finished all the write-ups and presentations, I would say that the project is completed. However, I would also argue that because of my slip up, the database could still be improved so that the extra ordered locus names are not counted.

Reflection on the Process

I honestly learned a lot from this project. With my head as a computer scientist, I have been exposed to creating and testing a database. Even though what we did here was just scratching the surface, it was still a worthwhile and enlightening experience. Learning about the Unix commands have actually boosted my knowledge in using the command line, and learning about another language like PSQL would later help me when I take a databases class as a senior. As far as biology is concerned, although most of the materials discussed in class went over my head, I would still say that my exposure to the biological aspect of databases would still be useful.

With my heart, I would say that my team worked really well. There were no problems that arose between any of us and we were all comfortable working together. I would also add that we communicated effectively with each other and there was no such occasion where we think that any of us are not doing their job as they are supposed to.

With my hands, I would definitely say that I learned how to use various tools in the Windows machine. From the command line to GenMAPP, my knowledge of these programs has gone up and I have become more familiar with them.

The lesson that I will take away from this project that I will still use a year from now is definitely using the command line such as grep and sed. Before this class, I never used them before and so I believe that in my next classes, they will be especially useful. A year from now, I will also be taking my databases class so I feel as though the knowledge I gained from using PSQL would aid me in that class.