Reflective Essay
BL layer
Author: Shahid Hadi
Colleagues: Georgina Leslie, DB layer, Vinod, FE layer.

Approach to the Project

At the very beginning, the name of the module was very tricky for me! I was expecting another course with the same format as Biocomputing_1. It took me some real time to understand the main aim of the module. Here, as a feedback, I would recommend the name of the module to be changed to Group Project rather than Biocomputing_2, as it is more straightforward! The online nature of the course was another real obstacle, as you have to work with people you have never met and you do not know anything about their perspectives and approaches regarding such a project, so it was real pain from the very start to get on the track!

The diversity of the subjects given in each session was quite hard to comprehend, especially for someone that comes from a biological background with little experience in programming. But I should confess that the takeaways of the module deserve every effort that you put in it, as it is the module that puts you directly at the heart of the bioinformatics field practically and teaches how to address the issues in the field and enables you to invest in it! Given all that, it was absolutely not easy to get on the track and give the project a start.

During our first meeting, we had a short introduction session, and I took the initiative divided the roles and assign a project manager. Both my colleagues had better experience than me regarding the requirements of the project. Upon agreement, Georgina took the DB layer, I took BL layer and Vinod took the HTML part. Because Georgina had a stronger programming background, we decided to make her the project manager. Both Vinod and Georgina had previous experience of HTML coding, Vinod decided to take the HTML part of the project. Georgina and I had previous experience, although mine was very limited, in logical BL layer, I wanted to invest more in the logical layer, as it is my favourite area of coding, I decided to take the BL layer and Georgina wanted to invest more in the database layer and develop her skills there, so, she took the DB layer.

Performance of the Development Cycle

Weekly zoom meetings were set. I created a WhatsApp group for instant connectivity. At the beginning, the development cycle was very slow. My colleagues face some extraordinary circumstances, and I had a long journey back home from London. It took us around three to four weeks to set our APIs. I and my DB colleague were using Github regularly, that is why our connectivity was more efficient. Our FE colleague preferred WhatsApp and emails for contact, this made our connectivity to be less efficient.

The Development Process

For integrity of the project, I asked the DB layer to return all the information that I need for my calculations at the BL layer in one single function. I developed my APIs and as soon as the APIs for the DB layer was ready, I started filling my API with real code. Most of my time went for design and iterations was used to refine the code and priming it to give the results in the best results possible. The first challenge was in handling the cds information, as it was the key information for all the calculations to be done at the BL layer. The cds indexing starts at 1 in the gene bank file, but indexing starts at zero in Python, also the slicing option in

python does not return the last index, so, both of these details should be counted for when handling the cds data returned by the DB layer.

Code Testing

I set a virtual environment in a JupyterLab, by using a dummy code for the function that contains all information that I need from the DB layer and filled the results of the function by real data of a gene from the gene bank file, so that I can work entirely independently and test every line of code before adding another line. This strategy was quite helpful, handy, independent and powerful. It helped to develop my code quite easily and add, remove or edit every bit of it quite easily, and bug fixing was quite easy. Every tested code was then pushed to Github and three versions of different stages of development files was saved on my PC.

Known Issues

There were issues, especially with the codon usage function, as the requirements were ambiguous and codon usage itself can have different meanings. But all of these issues were handled properly and solved efficiently.

Problems and Solutions

As I mentioned, codon usage function was quite a difficult task itself, added to it the complexity of the requirements and its ambiguity was quite challenging. But thanks to soft coding, the networking of my code depends on sophisticated indexing, I was able to come up with all the requirements and even more, I have provided a manual statistical test to show the significance of difference between the entire chromosome and a single gene. Without soft coding, editing the code would have been really difficult and new code should have been written instead! Not to forget the efforts of our FE colleague in understanding the exact requirements for codon usage function. His efforts lead to better insight of the task.

At the beginning, my function only returned the codon frequency, the ratio of a codon used for coding a certain amino acid to all the other codons used to code for that amino acid. But later, we edited the function to return the percentage of a codon in 100 codons used by a certain gene. Later, the statistical test was also added. All these are returned in a very sophisticated way linking every amino acid to all of its codon, every codon to its statistics, and every statistic number to its significance! It was quite a job! But led to nice results!

Alternative Strategies

My code for RE sites is somehow simple. But as we are doing a computer exercise, it is more than enough for such a purpose. Although my code is simple, as it does not count for the exact position that an enzyme cuts, instead, it searches for an entire match for the sequence recognized by the enzyme and its palindromic sequence, but the code skeleton is strong enough to hold this detail anytime needed. So, a more detailed code would have been possible, but for the purpose of our project, we have assumed that this would be enough

Personal Insights:

This project was extraordinarily useful. It puts me at the heart of the tools and resource used in the field of bioinformatics and made me learn through practice! The best way of learning ever! HTML, cgi scripts, design of interactive web pages, the way most of bioinformatic resources work, github version control, command line, and building a database open to a user through a web page! These are all the things that I learned from the project. What I learned from this module will help me very much for my MSc project and improve my employability to a good level! I t was quite hard! But, as I said at the very beginning, it deserves every efforts you put in it.