Advanced Practical Computer Concepts for Bioinformatics

Project Proposal Rough Draft – Matthew Jones

My idea for a practical bioinformatics program is an analysis tool that will accept either a DNA or amino acid sequence in FASTA format and report various data based on the query and store the information on the query in a database. The data reported may include: CG/AT composition, translated DNA, amino acid distribution, hydrophobic amino acid composition, atomic weight, and possibly more.

**The front-end**

I intend for the tool to begin with a form created with HTML, prompting the user to input either a DNA or amino acid sequence, either as pasted in a text box or uploaded as a file. This form will utilize JavaScript to validate user input and ensure that only DNA or amino acid data is processed. I intend to utilize other jQuery methods to further refine the form. Once the data is processed, it will be displayed graphically to the user using interactive CSS and HTML.

**The back-end**

This tool will be utilizing Python and its Biopython, MySQL Connector, and CGI modules to do the backend computing. The Biopython module comes with many useful functions and methods for parsing biological data, including DNA and amino acid sequences, and they will be utilized in this analysis tool. The CGI module will be used to communicate with the form and gather the data for Biopython to process. Using MySQL Connector, the resulting data will be used to populate a table in MySQL.