Advanced Practical Computer Concepts for Bioinformatics

Final Project Proposal

Claire Harelson – JHU ID: charels1

**Protein Structure Visualization and Analysis Tool**

Tool Overview

Structural biology is the study of the three-dimensional structures and dynamics of molecules. Protein structure analysis in bioinformatics is incredibly important as structure is directly related to function. Therefore, increased knowledge of a molecule’s structure can ultimately lead to enhanced understanding of its characteristics. Although a wide variety of structural bioinformatics tools exist, this particular tool will aim to provide a more user-friendly web interface to explore and analyze protein structures and related information. Users will be able to interact with the web application through a browser while data processing will occur server-side via CGI and JavaScript files as well as through interactions with a MySQL database.

Tool Structure:

* 1. **The UNIX OS and filesystem:** Initialize a project directory structure on the class server and appropriately organize files and corresponding folder system.
  2. **Placement and organization of files within a web application:** Within the class server, create applicable directories for HTML, CSS, JavaScript, and CGI files. Additionally, store protein structure files in PDB format in an allocated folder.
  3. **Python and CGI programming:** CGI scripts will implement Python logic to handle HTML form requests and responses to maintain content separation. Server-side functionality will be implemented here to process user inputs. Python will be used to parse information from PDB files into a MySQL database.
  4. **CGI and HTML templates:** CGI scripts will interact with HTML templates in order to continue maintaining appropriate content separation. HTML templates will be utilized to display the relevant protein information, which may include 3D structure visualization, results of a desired analysis, and so on.
  5. **HTML5 document markup:** HTML5 markup will be used to create a form that allows a user to input a query. For example, this may be to search a particular protein and return its relevant information. It will also be implemented to structure the web page properly and interact with the relevant CSS, CGI, and JavaScript forms. (Potential use of HTML canvas for 3D structure visualization if I’m able to figure that out.)
  6. **Relational database schema design:** Protein-related data will be stored in a relational database, including information about their structures, functions, and other properties.
  7. **MySQL:** A MySQL database will need to be initialized to store and manage the structural data from the PDB files.
  8. **Using mysql.connector python module to connect to MySQL:** The msql.connector python module will be used to establish a connection between the MySQL database and web application.
  9. **Limited page styling with CSS:** Utilize CSS to create an aesthetic layout for the web page. Separate style page will be used to maintain content separation.
  10. **Javascript and JQuery for client-side interaction:** Javascript and JQuery will be used to add client-side interaction to the web application and interact with the HTML page. Potential use of Java libraries like Three.js or BioJS to visualize 3D structure.

**Data generation:**

**-**Based on query input

-Searches PDB files via MySQL

-Returns results to web page through CGI scripts

**Query output**

**User input**

**MySQL query output with resulting PDB information**

**User input:**

-Name of protein of interest to search and desired associated information

**MySQL query to database with PDB information**

**Query results: (CSS/HTML/JavaScript)**

-Name of protein

-Summary of protein

-Method of structure determination

-Protein structure

**PDB Information**

**Database**

**Python CGI**

**User**

**Web Page**