Chris Dimapasok

Advanced Practical Computer Concepts Project Proposal

**Background**

The basis of this project is to generate a streamlined pipeline that will extract genetic information, and store biologically relevant information that will have a role in understanding disease. More specifically, the project will utilize a gene ontology/ gene functionality(annotation) tool to query for gene information related to a disease and extract gene annotation results and their molecular function. Additionally, we will implement a drug treatment tool to identify possible drug targets/treatments that may have a chance at treating or improving a disease. Since there aren’t many sources that combine both of the information together, we hope to create concise tables with both treatment options and genetic information into a singular webpage. The webpage will be a simple, user-friendly page that will be robust with genomic annotation information and treatments that is filtered to only the elements of interest to the user. This will enable patients and researchers to be informed of the treatment possibilities that patients can look at and consider, as well as further understand the genes involved in disease.

**Pipeline**

The gene ontology predictor tool Blast2GO will first be used to generate data including the gene name, function, annotation information, and more. The tool requires FASTA formatted files that contain genomic sequences that code for a protein. To execute the first step in the pipeline, we will download the required FASTA file from the NCBI website. Using the different Blast2GO features like blast and interproscan will generate information that can be stored at our discretion. We will download the data and parse only the relevant information that we want to include in our analysis, such as the annotation information, gene name/id, and molecular function.

Additionally, we will use a drug target interaction tool to search for drug targets related to a disease. The web application, called DT-Web uses a computational algorithm to identify possible drug combinations and recommendations for patients. The parameters of interest include the therapeutic indication (treatment) and drug name. We hope to combine both the gene ontology information and the drug information in a sql database schema. Here, the database will contain relevant functional genetic protein and annotation information related to the disease as well as possible drug targets and treatment options.

Once the database schema is created, the user will input a search term, and the CGI will process this and query from the SQL database based on the information entered from the user, using a SQL connection. The jinja2 module will be utilized for webpage functionality, and a html template will be created. As the user inputs information, the Jquery autocomplete feature will be implemented to suggest similar search terms to whatever the user is entering. For example, if a user searches for part of a gene, it will autocomplete with genes that are closely related to the search term. If a user is attempting to search for a gene product or drug, the filter will automatically complete for various gene products/drugs.

Once entered, the CGI will process the information, search through the sql database, and place the information in stylized tables. One table will contain the significant genetic information from the sql database that the user searched for and the other table the different treatment options available. We will use CSS to add style and additional functionality to our tables, specifically highlighting significant elements that may be important for the user.