**FASTA Manipulation Tool – Project Proposal**

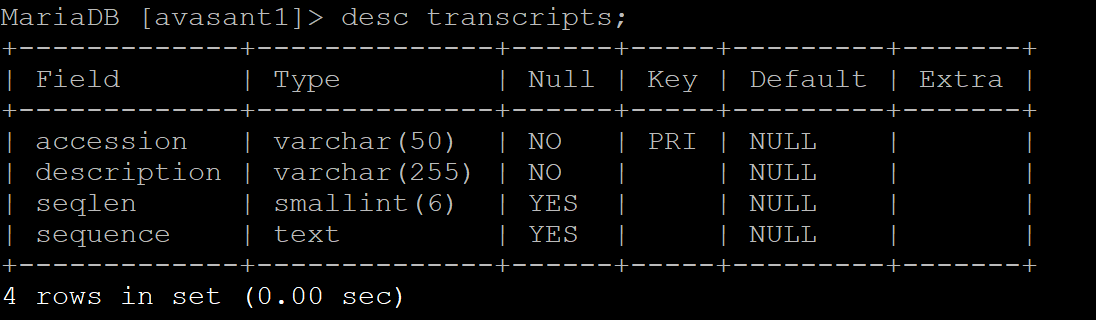
**Introduction:**

The FASTA manipulation tool provides an interface for the user to select a FASTA sequence to investigate. The sequence header with the accession ID, description, and sequence length are displayed. Users can click on the hyperlinks provided for these records which would direct them to the relevant page at NCBI. A search box feature will be provided to inspect the FASTA files in the repository. The FASTA sequences that match the search criteria will be displayed in the interface. A checkbox will be provided beside each record giving the user a choice to save the sequence information into a table in MYSQL database. Users will also have an option to retrieve the records that are saved.

**Technical specifications:**

The FASTA files will be retrieved from NBCI [1]. These would correspond to human genes that are most studied due to their clinical significance. The files will be deposited in a directory in the bfx3 server. In the first run of the tool, the interface would offer the users a dropdown list with the list of files in the repository. Once the user submits the choice, the CGI script will read the input choice and retrieve the corresponding FASTA file from the server directory. The script will then parse the FASTA file, retrieve the ID, description, and sequence length. This would be displayed in the interface along with a search box. The ID field of the record displayed would be hyperlinked to the associated record at NCBI. The user can then inspect the transcript information in detail associated with the NCBI record. The search box made visible using JavaScript and CSS would allow users to search by terms in the FASTA header. Upon submission of the search term, the CGI script will look for the term in all the FASTA files in the directory. The files with matching terms will be retrieved, their information parsed and displayed back in the interface. Each record will have an associated checkbox which will give the user the option to save the FASTA information to a table in MYSQL database. There will also be an option to retrieve the records that are saved in the table based on the description field and display them in the interface.

**Database**: The database ‘avasant1’ will be used in the program. A table called ‘transcripts’ will be used to save the records. The fields in the table are accession, description, length, and sequence with accession being the primary key. The structure of the transcripts table can be seen in the screen shot below:



References:

[1]. https://www.ncbi.nlm.nih.gov/search/