Data Foundation Systems

Team DFASTA
Software Requirements Specification

Overview

Team DFASTA aims to build a Genome Browser which will primarily support FASTA and FASTQ formatted genome data. The browser should allow users to browse through stored FASTA files in an efficient manner and allow for comparisons between two genomes through a variety of sequencing algorithms. The browser will also support anonymization of genome data.

System Requirements

Functional Requirements
- The genome browser shall support the following types of genome sequences:
    - FASTA files
    - FASTQ files
- Users can:
    - Compare two genome sequences side by side and get a score using multiple comparison algorithms such as:
        - Smith-Waterman
        - Needleman-Wunsch
        - HMM - Hidden Markov Model
    - Compare input genome against NCBI's BLAST database
- The system will offer the ability to anonymize genome sequences, allowing selective anonymization based on various genome-related parameters
- Users should have the ability to:
    - Perform text-based searches within FASTA files

Non-Functional Requirements

- Performance
  - The system should provide efficient response times, even with large genome sequences.
- Security
  - User data and genome sequences must be securely stored and accessed.
  - Anonymization must comply with privacy and security regulations.
- Compatibility
  - The genome browser should be compatible with a range of modern web browsers and devices

Project Deliverables

- Fully Functioning Web Application
  - Frontend
    - ReactJS
  - Backend
    - Python - Flask
  - Database
    - MongoDB