

Kevin Feng Yong Chen
JHU Email: fchen29@jh.edu
Student ID: FE4A8B

Program Description

This program takes sequences from an input file and outputs the length and the sequence of the longest common subsequences (LCS) between the various pairs of sequences from the input file.

This program was written in python3 and was developed using the Spyder IDE.

Please follow the following format for the input file:

S1={required sequence 1}
S2={required sequence 2}
S3={required sequence 3}
S4={required sequence 4}

{test case 1}
{test case 2}
{test case 3}

.
.
.

Please make sure that the 4 required sequences are in the first 4 lines of the input file and make sure that they are prefixed with “S1/2/3/4 =” so that the program can recognize them as required sequences. If the program does find 4 required sequences, it will treat the situation as an error and self terminate. The test cases follow the required sequences. Empty lines are ignored by the program.

The output follows the following format:

comparing between the required input strings:

LCS between S1 and S2:

{sequence of LCS}
{relevant statistics}

LCS between S1 and S3:

{sequence of LCS}
{relevant statistics}

.
.

comparing the test cases to each of the required input strings:

test case 1: {test case sequence 1}

S1: {S1 sequence}
LCS: {sequence of LCS}
{relevant statistics}

.

.

.

Prior to printing the results, the 4 required sequences from the input file are printed to the top of the output file.

Usage

Please follow the following format when executing the script through command line:

Python3 main.py -i <input_file> -o <output_file>

The path to the input and output file must be provided. In addition, the program provides the -h option that tells the user how to run the program.

Input

The script accepts text files as input files.