BLG 336E - ANALYSIS OF ALGORTITHM 2

HOMEWORK 3 – REPORT

1.

a. In the main function, strings file, output file, match, mismatch and indel score are taken as main arguments. Two vector of strings defined. Read the strings.txt file and push all of them into str vector. Than this vector is sorted alphabetically. Then for each combination, sweat function is called. Sweat function returns a vector. In the sweat function, the sequence operation, matrix filling, traceback are made. Firstly, fill the matrix full of zeros. After that, the maximum value of the left side, upper side, and the diagonal side are calculated, and whichever is the largest is taken as the maximum value. If it is a negative value, 0 is taken. After the matrix fullfilled with thoses values, then traceback operations start. Starting from maximum value, move forward until you see 0. While doing that, put the chars into a vector of chars. Thus, at the end, take the string from these vector of chars. Sort them again and return vector of strings to a main function. Then it will be written to output.txt file.

b. The Smith- Waterman algorithm has O(nm) complexity because of the matrix M[n][m]. In our case, the complexity is O((str1.length).(str2.length.)). Because for each element we make calculations. These complextiy is valid for just two strings. But in our case, there is a lot of strings. So, in our case total complexity is $\binom{num\ of\ strings}{2}$xO((str1.length).(str2.length.)).

2.

a. For example in the brute force approach, more calculations should be made. Because it calculates all possibilities one by one and this is time consuming and have more calculations. In the Smith-waterman algorithm, create a matrix and find from these matrix most common sequence. It is more practical.

b. Again for example, in the brute force there may be find 3 possible sequence alignment. Smith waterman finds less. Thus, the calculation results kept in the memory is less in the Smith-waterman algorithm.

c. From the above explanations, the running time of the smith waterman is shorter as brute force requires more calculations and more processing.