CMPS 102 — Fall 2018 — Homework 4

"I have read and agree to the collaboration policy." - ***Kevin Wang***

Collaborators: *None*

Solution to Problem 2: DNA

2 DNA sequences $S = \{s_1, s_2, \cdots, s_i, \cdots, s_m\}$ and $T = \{t_1, t_2, \cdots, t_j, \cdots, t_n\}$ have lengths m and n respectively.

Let LCS be the longest common sub-sequence.

LCS$[i, j]$ is the length of the largest common sub-sequence, LCS, shared by two sequences of sizes i and j.

Base Case: An empty sequence does not share a common sub-sequence.

$$LCS[i, 0] = 0 \text{ and } LCS[0, j] = 0$$

Case 1: $s_i = t_j$; the element is a part of the LCS.

$$LCS[i, j] = 1 + LCS[i - 1, j - 1]$$

Case 2: $s_i \neq t_j$; we ignore one of the elements.

$$LCS[i, j] = \max(\ LCS[i - 1, j]\ ,\ LCS[i, j - 1]\ )$$

Therefore, the LCS of S and T has length $LCS[m][n]$. We can then walk back through the matrix to find an LCS.

Continued...

Algorithm 1 Finds the LCS of of two sequences S and T with sizes m and n, respectively

LCS (m, n):
Let $LCS[m][n]$ be the sub-sequence size matrix
for $i = 0$ to m **do**
  for $j = 0$ to n **do**
    if $i = 0$ or $j = 0$ **then**
      $LCS[i][j] = 0$
    else if $s_i = t_j$ **then**
      $LCS[i][j] = LCS[i-1][j-1] + 1$
    else
      $LCS[i][j] = \max(\ LCS[i-1][j]\ ,\ LCS[i][j-1]\ )$
    end if
  end for
end for
$LCS[m][n]$ is the length of the LCS

Let LCS be an empty sub-sequence
Let $i = m$ and $j = n$
while $i > 0$ and $j > 0$ **do**
  if $s_i = t_j$ **then**
    $s_i :: LCS$
    $i--, j--$
  else if $LCS[i-1][j] \geq LCS[i][j-1]$ **then**
    Using $LCS[i-1][j] > LCS[i][j-1]$ can return another different LCS
    $i--$
  else
    $j--$
  end if
end while

Time Complexity: $O(mn)$
The 2 for-loops iterate m and n times, respectively, taking time $O(mn)$. Iterating back through the matrix to find the elements of the LCS takes $O(m + n)$. Total time used is $O(mn + m + n)$.

Space Complexity: $O(mn)$
The LCS matrix takes size $O(mn)$. The LCS sequence, takes (at most) $O(\max(m, n))$. Total space used is $O(mn + \max(m, n))$.