# Assignment 4: Sequence Alignment

## Learning Outcomes

* Learn a dynamic programming approach to a difficult problem.

## Motivation

The book has a section devoted to Sequence Alignment, which is similar to the Levenshtein distance discussed in class.

All of this is described in detail in the book.

## Inputs

This program will accept strings each entered one at a time. The strings may be different lengths.

* Each string will represent a sequence of genetic code and consist only of the letters A, C, G, and T.
* There will be no spaces or other symbols.
* Each string will be on a different line.

## Outputs

Your program should report the optimal alignment cost in order to align the two strings.

(This line is from the book.) For the sake of concreteness, we assume the following:

* The penalty for a mismatch is 1.
* The penalty for a gap is 2

## How to solve the problem.

Here are the steps followed by the book.

1. Create an matrix, where is the length of the first string and is the length of the second string.
2. Set the bottom right element of the matrix to 0.
3. Starting from the bottom right and working up, set each element to be 2 plus the element that is below it. The bottom most element is 0.
4. Starting from the bottom right and working left, set each element to be 2 plus the element that is to the right it. The right most element is 0.
5. Now, repeat the following with a double-for loop. The outer for loop (named "j") should span from to 0 and work backwards in decrements of 1. The inner for loop (named "i") should span from to 0 and work backwards in decrements of 1.
   * If the character at first at position is equal to second at position , copy the value at into the matrix at position .
   * Else, do the following:
     + Fetch the element below this position and add 2 to it.
     + Fetch the element to the right of this position and add 2 to it.
     + Fetch the element below AND to the right of this position and add 1 to it.
     + Find the smallest of the three fetched values. Write that value to the matrix at position .
6. Return the element at position . This should be the score of the aligned sequence.

## Examples.

The book contains one example.

Enter first sequence: AACAGTTACC  
Enter second sequence: TAAGGTCA  
Optimal Alignment Distance: 7

Here's a problem from the chapter at the end.

Enter first sequence: CCGGGTTACCA  
Enter second sequence: GGAGTTCA  
Optimal Alignment Distance: 8

## Turn it in

Upload a zip file containing the files that use used or wrote yourself to the drop box on D2L:

* Make sure your name, CSCI 4270, and the Programming Assignment Number appear in comments in all of your files.
* Note: NO CREDIT will be given to programming assignments that do not compile.
* Make sure you have compiled and tested your program before handing it in.