Homework 3
CS 5046 (Spring 2012)

Assigned on March 19, Due by 4pm, March 28, 2012
Submit by email to baguilar AT cs DOT vt DOT edu

Problem 1

(4 points) In class we have implemented a script called distance.pl (available in the Web page ) with a Dynamic Programming algorithm to find the Edit distance between two sequences S1 and S2. Basically, the Perl script computes the matrix \( D(i,j) \) which represent the edit distance between the substrings \( S1[1..j] \) and \( S2[1..j] \). The edit distance between the strings \( S1 \) and \( S2 \) is found in the last element of this matrix (\( D(n,m) \)). Modify the script distance.pl to make it print an alignment (just one solution) of \( S1 \) and \( S2 \) from the matrix \( D \). Use \( S1=GTCACT \), and \( S2=GGATCC \).

Problem 2

(4 points) Implement a script called wdistance.pl with a Dynamic Programing algorithm to find the General Weight for edit distances. In this problem, which is an extension of the original Edit Distance problem, a weight is associated to some operations(mutations): a weight \( ws \) for insertion or deletion, and a weight \( W(r,s) \) for replacing \( r \) with \( s \). \( r \) and \( s \) can be A, C, G, or T. The dynamical program algorithm computes the matrix \( D(i,j) \), the weight of the optimal alignment of the substrings \( S1[1..i] \) and \( S2[1..j] \).

The Base case is : \( D(i,0)=i*ws \) and \( D(0,j)=j*ws \)

And the General Case (recurrence relation) is

\[
D(i,j)=\min \left( D(i-1,j-1) + W(S1(i),S2(j)) , \\
D(i-1,j) + ws , \\
D(i,j-1) + ws \right)
\]

Implement the previous algorithm and display the solution (\( D \rightarrow [n,m] \), \( n \) is the size of \( S1 \) and \( m \) is the size of \( S2 \)) for \( S1=GTCACT, S2=GGATCC \), using \( ws=3 \) and \( W \) equal to the following table:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Submitting your Homework

Create a directory called \(<YourName>-Homework3\) and put distance.pl and wdistance.pl in that directory. Submit your homework by zipping (or tarring and gzipping) the directory and emailing the zipped file to me.