PS5 DNA Sequence Alignment

This task required us to compare two ASCII strings in order to determine their edit distance, as well as provide space and time analysis on the program while it was running. Using a dynamic programming method, we were required to create a program that computed an optimum sequence alignment on two DNA sequences. The space complexity of our program might then be measured using Valgrind, a programming tool.

Key concepts

The EDistance class represents a matrix of integers with a vector of vectors of type int. The function optDistance() would then use the min() and penalty() functions to fill the matrix from bottom to top with the smallest of three distances, then return the optimal distance between the two sequences. The alignment() method would then go backwards through the populated matrix, constructing the output alignment string depending on the best path.

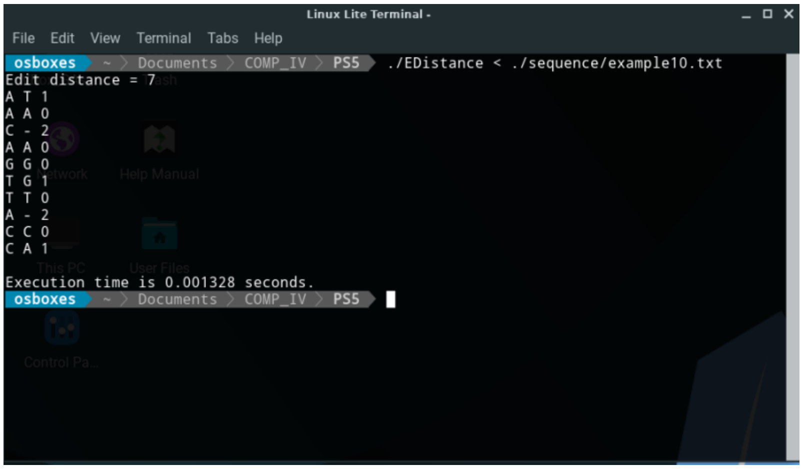
What I Accomplished

My EDistance algorithm uses dynamic programming to fill a matrix with each computation from bottom right to top left, culminating the ideal distance at [0][0]. Tracing the matrix in reverse order from top left to bottom right also shows the alignment path. Instead of computing the same subproblem numerous times, the dynamic programming technique allows the alignment computations to be divided down into subproblems and then stored.

What I Learned

Apart from the numerous applications of Edit Distance, I learned about the advantages of dynamic programming, which, when compared to the recursive solution for this assignment, would have had a much higher space complexity due to the number of recursive calls exceeding 2N when comparing two strings of length N. In addition to, I also gained the knowledge on how to use the Valgrind massif tool to efficiently monitor program memory use. This utility shows how much memory the heap consumed during execution. I was able to determine the projected run time and memory use of a bigger sample of strings of length N using the doubling approach.

Output



```
osboxes  ~ 〉Documents 〉COMP_IV 〉PS5  ./EDistance < ./sequence/example10.txt
Edit distance = 7
A T 1
A A 0
C - 2
A A 0
G G 0
T G 1
T T 0
A - 2
C C 0
C A 1

Execution time is 0.001328 seconds.
osboxes  ~ 〉Documents 〉COMP_IV 〉PS5 
```