```
/**********************************************************************
 *  readme
 *  DNA Sequence Alignment
 **********************************************************************/
```

Name: Matthew Lorette Anaya

Hours to complete assignment: 5

```
/**********************************************************************
 * Explain which approach you decided to use when implementing
 * (either recursive with memoization, recursive without memoization,
 * dynamic programming or Hirschberg\222s algorithm). Also describe why
 * you chose this approach and what its pros and cons are.
 **********************************************************************/
```

Implementation of this program was done with the use of dynamic programming and a matrix. I used the algorithm on the Princeton site in order to fill said matrix. In-order to find the alignment I used backtracking top-left to bottom right, moving from the current matrix index to the next-lowest matrix index. There was a certain case where if the diagonal was 1 higher than the downwards or rightwards option, diagonal was still the taken rout. In any case, depending on which direction I went, I either added a gap, or both letters, and incremented i and j counters to traverse back to the bottom right of the matrix.

```
/**********************************************************************
 * Does your code work correctly with the endgaps7.txt test file?
 *
 * This example should require you to insert a gap at the beginning
 * of the Y string and the end of the X string.
 **********************************************************************/
```

Kinda confused here on what this question really is. The pdf is using example10.txt
And this is asking for endgaps7.txt. Seems like there is a mix up of pdfs between different years of this . So I’m going to use the what the HW pdf says as there really isn’t an example to compare to
otherwise. Though it also says to put this all into a folder named ps3, which is definitely incorrect.

Input:

```
        Ê./EDistance < ./sequence/example10.txt
```

Expected output:

```
        Edit distance = 7
        AT1
        AA0
        C-2
        AA0
        GG0
        TG1
        TT0
        A-2
        CC0
        CA1
```

```

What happened:

        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 0.00094 seconds.




/**********************************************************************
 * Look at your computer\222s specs in the settings.
 * How much RAM does your computer have and explain what this means?
 **********************************************************************/

My Mac has 16gb of RAM. Random access memory gives applications a place to
store and access data on a short-term basis. It stores the information your
computer is actively using so that it can be accessed quickly.


/**********************************************************************
 *  For this question assume M=N. Look at your code and determine
 *  approximately how much memory it uses in bytes, as a function of
 *  N. Give an answer of the form a * N^b for some constants a
 *  and b, where b is an integer. Note chars are 2 bytes long, and
 *  ints are 4 bytes long.
 *
 *  Provide a brief explanation.
 *
 *  What is the largest N that your program can handle if it is
 *  limited to 8GB (billion bytes) of memory?
 **********************************************************************/

N^2 is the area of the matrix, the number of integer slots that need to be
filled in. 4 is the size of an integer in bytes.

a = 4
b = 2
largest N = ~44,721

Explanation:
        4 * 44,721^2 = 7,999,871,364 just shy of 8gb.

/**********************************************************************
 * Run valgrind if you can and attach the output file to your submission.
 * If you cannot run it, explain why, and list all errors you\222re seeing.
 * If you can run it successfully, does the memory usage nearly match that
 * found in the question above?
 * Explain why or why not.
```

```
/**********************************************************************

-------------------------------------------------------------------------------
  n       time(i)         total(B)   useful-heap(B) extra-heap(B)    stacks(B)
--------------------------------------------------------------------------------
 67  6,808,807,498    3,085,122,584    3,084,577,402       545,182            0
 68  6,870,407,882    3,146,681,624    3,146,125,562       556,062            0
 69 30,570,148,197    3,201,904,240    3,201,338,395       565,845            0

It does not, its actually quite different and I’m not entirely sure as to why. Not
Sure if I’m reading valgrind output wrong or my equation is.


/**********************************************************************
 *  For each data file, fill in the edit distance computed by your
 *  program and the amount of time it takes to compute it.
 *
 *  If you get segmentation fault when allocating memory for the last
 *  two test cases (N=20000 and N=28284), note this, and skip filling
 *  out the last rows of the table.
 **********************************************************************/

data file           distance       time (seconds)
-------------------------------------------------
ecoli2500.txt        118                       0.125216
ecoli5000.txt        160                       0.334861
ecoli7000.txt        194                       0.521017
ecoli10000.txt       223                1.4272
ecoli20000.txt       3135              74.6052
ecoli28284.txt       8394             177.645

/************************************************************************
 *  Here are sample outputs from a run on a different machine for
 *  comparison.
 ************************************************************************/

data file           distance       time (seconds)
------------------------------------------------
ecoli2500.txt          118            0.171
ecoli5000.txt          160            0.529
ecoli7000.txt          194            0.990
ecoli10000.txt         223            1.972
ecoli20000.txt         3135           7.730



/**********************************************************************
 *  For this question assume M=N (which is true for the sample files
 *  above). By applying the doubling method to the data points that you
 *  obtained, estimate the running time of your program in seconds as a
 *  polynomial function a * N^b of N, where b is an integer.
 *  (If your data seems not to work, describe what went wrong and use
 *  the sample data instead.)
 *
 *  Provide a brief justification/explanation of how you applied the
 *  doubling method.
 *
 *  What is the largest N your program can handle if it is limited to 1
 *  day of computation? Assume you have as much main memory as you need.
 **********************************************************************/
a =
b =
largest N =

/
**********************************************************************
 *  Did you use the lambda expression in your assignment? If yes, where
 * (describe a method or provide a lines numbers)
 **********************************************************************/
No



**********************************************************************
 *  List whatever help (if any) you received from the course TAs,
 *  instructor, classmates, or anyone else.
 **********************************************************************/

N/a

/**********************************************************************
 *  Describe any serious problems you encountered.
 **********************************************************************/



/**********************************************************************
 *  List any other comments here.
 **********************************************************************/

```