**ECE 1111: Engineering Computation I**

**Laboratory No. 10: Windowed Grep in C and Python**

**Deposit your work in:**

**/data/courses/ece\_1111/current/labs/lab\_10/<lastname\_firstname>**

**Goal:** Locating text in a file is one of those extremely fundamental capabilities that Unix excels at with commands such as grep. In this assignment, you will write a simple version of grep that provides some flexibility for how the output is presented.

**Description:**

(1) Write a Python program that has the following user interface:

**mygrep.py -w “word” -n num\_lines file1.txt file2.txt ...**

The program cycles through all the files specified on the command line and searches each line for a match to *“word”*. The match should be **case-insensitive**. The options “-**w**” and “**-n**” can be specified in any order. The option “-**w**” is mandatory – if it is not specified the program should display an error message and exist. The “**-n**” option is optional with a default value of $1$.

When you find a match, output the lines immediately before and after the match. For example, suppose you specify the options *“***-n 2 -w bOB foo.txt***”* and the file **foo.txt** contains:

**See Jane run.**

**See Mike run.**

**See Mary run.**

**See Bob run.**

**See Alice run.**

**See Tom run.**

**See Barbara run.**

The output should be:

**===**

**File: foo.txt**

**See Mike run.**

**See Mary run.**

**SEE BOB RUN.**

**See Alice run.**

**See Tom run.**

**===**

In other words, you should print the filename that produced the match, line that matched and the preceding two lines and following two lines because *“*$n = 2$*”*. Print the line that matched in upper case and make sure the match is case-insensitive (note the mixed case in my example).

Do not read the entire file into memory. Read the file line by line so that you can process arbitrarily large files.

(2) Repeat (1) in C. For the C implementation, think about how you can use an array of character string pointers as a data structure to facilitate maintaining the most recent $2n+1$ lines from the file in memory. Create a structure called Buffer to hold the data. This structure should have an array of character string pointers to hold the $2n+1$ lines of text, the value of n and the size of the window ($2n+1$).

You can use **fgets** to read the file, but you must use a streaming data approach (meaning buffering the data in a buffer of $2n+1$ lines).

Time your code on a set of text files that have one million lines of text in them. Compare the execution time of your Python code, your C code and the Unix version of grep. Put your results in the top of your Python file as comments.