CSE 5311 Lab Assignment 3

Due August 4, 2008

Goals:

1. Understanding of suffix arrays, along with the related notions of suffix rank and longest common prefix (lcp).

2. Application of the above concepts to the ***longest common substring*** problem.

Requirements:

1. Write (and test) a C/C++ program to determine a longest common substring of two input strings with no more than 500 symbols each. Each string will be on a line by itself. These strings will be readable using `scanf` or streams, since they will only include upper/lowercase alphabetic symbols or digits. Your code should run in time $O(n)$, not including the time to compute the suffix array, where *n* is the total number of input symbols.

   Your output is any maximum length string common to both input strings. It is important to find a ***substring*** (i.e. contiguous symbols) rather than a ***subsequence***.

   Your program must compile and execute on OMEGA. There should be a comment near the beginning of your code that indicates how to compile on OMEGA. Your debugging trace should be disabled in the version you submit.

2. Email your code (as attachments) to `mazloom@uta.edu` before 10:15 am on August 4. The subject should include your name as recorded by the University.

Getting Started:

1. You may use suffix array code from the course webpage or elsewhere.

2. Symbol-to-symbol comparisons should occur only during the preprocessing before finding the longest common substring.