Homework 6

Out: 11.2.16
Due: 11.14.16

1. [String matching, 20 Points]
   a. Draw the FSM to find the pattern "abac". Assume that $\Sigma=\{a,b,c\}$, and that any letter not in Σ returns to the initial state.
   b. Draw a standard trie for the following set of strings:
      {abab, baba, ccccc, bbaaaa, caa, bbaacc, cbcc, cbca}.
   c. Draw a compressed trie for the set of strings from part (b).

2. [Graph representation 15 Points]
   a. Graph G is an undirected graph, represented using the following adjacency list. Draw G.

Vertex	Neighbors
1	2,3,4
2	1,3,4
3	1,2,4
4	1,2,3,6
5	6,7,8
6	4,5,7
7	5,6,8
8	5,7

   b. Which graph representation would you use to represent a graph with V vertices and E edges if E is O(V) for minimum space requirement? What if E is $O(V^2)$? What if we also care about adjacency time?
   c. A universal sink is a vertex with in-degree |V|-1 and out-degree 0. Given a graph G(V,E), represented as an adjacency matrix, describe an algorithm that determines whether G(V,E) contains a universal sink.

3. [Graphs 25 Points]
   For which integer n > 0 values do the following graphs have the given chromatic number? Explain.
   a. C_n, chromatic number 2.
   b. S_{n+3}, chromatic number 2.
   c. W_n, chromatic number 3.
   d. Q_n, chromatic number 3.
   e. $K_{n,n}$, chromatic number 3.

4. [Big Data, 40 points]
   This problem involves finding information in a very large data file, *BigData.txt*, similar to what you might find from the dump of a hard drive. The file can be obtained from the /ad/eng/courses/ec/ec330/BigData, or from the provided link on

Blackboard. Please let us know if you have trouble accessing the file.

Within this file, determine:

a. The number of "words" that contain only letters. Here, we define a word as a sequence of characters delimited a white space (that is, any number of spaces, new lines, tabs, etc).

b. The shortest pangram you can find within the text. A pangram is a sentence using every letter of the alphabet at least once. Any letter could be in either lower case or upper case, and the sentence may include any non-letter characters.

c. The number of English words (from the *dictionary.txt* file, case sensitive) that appear in the file that do not contain the first letter of your first name (either in lower-case or capital letter). The words need to be delimited by spaces, or other punctuation (i.e. anything that is not a number or a letter). Duplicates count. For example, the word "a", which is a dictionary word, appears in the data more than once, and should be counted accordingly.

Provide your solutions (i.e. your output to each of the sections), along with a brief explanation of each of your solutions and their runtime, in a file named *Problem4.txt*. Submit this file, along with all your solution code on Blackboard. Your code must be written in C++ and have three functions, *fourA*, *fourB*, and *fourC*, that runs with no arguments in the same directory as *BigData.txt*. Your code should print the results that you report in *Problem4.txt* to the screen.