

**Homework 6****Out:** 11.3.20**Due:** 11.11.20**1. [String matching, 15 Points]**

- Draw the FSM to find the pattern “aabaaa”. Assume that  $\Sigma=\{a,b\}$ , and that any letter not in  $\Sigma$  returns to the initial state.
- Draw a standard trie for the following set of strings:  
{day, dad, dill, did, so, soap, son}.
- Draw a compressed trie for the set of strings from part (b).

**2. [Strings, 5 points]**

Design an efficient algorithm that determines if a string  $S$  is a rotation of another string,  $S'$ . For example, ‘cdeab’ is a rotation of ‘abcde’. Provide the runtime of your algorithm.

**3. [Graphs 20 Points]**

For which integer  $n > 0$  values do the following graphs have the given chromatic number? Explain.

- $C_n$ , chromatic number 2.
- $S_{n+3}$ , chromatic number 2.
- $W_n$ , chromatic number 3.
- $Q_n$ , chromatic number 3.
- $K_{n,n}$ , chromatic number 3.

**4. [Graph representation, 10 Points]**

- 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

- 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?

**5. [Big Data, 50 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.txt or

<https://drive.google.com/file/d/0B2H-ZCNPqkX2bzhxUIRVdHg4cVE/view?usp=sharing>

Let us know if you have trouble accessing the file.

Within this file, determine:

- a. The number of BU-IDs in the file whose digits add to a number between 31 and 68 (including 31 and 68). A BU ID is defined as anything starting with a U and followed by exactly eight digits and then a non-digit.
- b. The number of English words (from the *dictionary.txt* file) that appear in the file that do not begin with the same letter of the alphabet as your first name (for example, if your first name begins with the letter 't', then you should exclude dictionary words that begin with 't'). Note that case matters (e.g. 'Ab' is a dictionary word, and it is not considered the same as 'ab'). The words do not need to be delimited by spaces (or other punctuation) in the file and may overlap. The words must be contiguous e.g: door\$knob or door knob will not find doorknob, but will still find door, knob, or, etc. Duplicates do count. For example, the word "a", which is a dictionary word, appears in the data more than once, and should be counted accordingly.
- c. The longest palindrome (i.e. a string that equals its reverse) you can find within the text. When finding palindromes, consider ALL characters (spaces should count).

Provide your solutions (i.e. your output to each of the sections, along with a brief explanation of each of your solutions), in a file named *Problem5.txt*. Submit this file, along with your solution code. Your code must be submitted in a single file named *Problem5.cpp*. Your code should include three functions, *fourA*, *fourB*, and *fourC*, that when run with no arguments in the same directory as *BigData.txt* print out the results that you report. In order to receive full credit, your solution needs to be efficient.