Assignment 2
Due Friday February 15th, 2019

Submission Instructions: Submit solutions in a single PDF via OWL. Assignments are due at 11:59:59 pm (Eastern Time) on the date listed above. As per the course late policy, assignments submitted more than 48 hours late will **not be accepted** and a mark of zero (0) will be recorded. Email submissions will not be accepted.

1. [12 marks] Regular Expressions

(a) [4 marks] Construct an ε-NFA that represents the language defined by the following regular expression:
$$R = 10^*(01)^+ + 0(00 + 11)^*$$

Note the '+' notation means "or", e.g., $(00 + 11)$ is a string consisting either of 00 or 11.

(b) [4 marks] Construct an ε-NFA that represents the language defined by the following regular expression:
$$R = 1^*00^* + (00 + 1)^*$$

Note the '+' notation means "or" notation $(0 + 1)$ means 0 "or" 1.

(c) [4 marks] On the alphabet $\Sigma = \{0, 1\}$, if we assume that 0 and 1 are binary digits and the strings in our language represent numeric values, describe, in English, the language that the following regular expression represents:

$$R = (1 + 0)^*00^+ + (100 + 1100)^*$$

2. [18 marks] Pumping Lemma

(a) [6 marks] Use the pumping lemma to prove language L_a is not regular:

$$L_a = \{1^n 2^n 3^n \ : \ n > 0\}$$

(b) [6 marks] Use the pumping lemma to prove language L_b is not regular:

$$L_b = \{1^a 2^b \ : \ 0 < a < b\}$$

(c) [6 marks] Use the pumping lemma to prove language L_c is not regular:

$$L_c = \{a^p \ : \ \text{p is prime, i.e.} \ p = 2, 3, 5 \ldots\}$$

3. [12 marks] Context-free Grammars

For each of the following languages, give the associated context-free grammar:

(a) [3 marks]

$$L_a = \{ww^R \ : \ w \in \{a, b, c\}^*\}$$

Note: w^R denotes the reverse of a string w.

(b) [3 marks]

$$L_b = \{0^*101^* : \text{the number of 0s is even and the number of 1s is odd}\}$$

(c) [3 marks]

$$L_c = ((011^* + 1^*) + 1110)^*$$

(d) [3 marks]

$$L_d = \{0^i 1^{i+j} 2^j \mid i, j > 1\}$$

4. [8 marks] Regex in Practice

For this question you will perform string matching on a large text file. Download the following text file containing the entire collected works of Shakespeare:

http://www.gutenberg.org/files/100/100-0.txt

Use the Unix/Bash command-line utilities like grep, awk, uniq and sort. If you don't have access to a Bash command-line, you can either install an emulated version in Windows, or try one of the these many pre-configured Linux virtual machine images.

Tip: To download the file directly from the command line try:
`wget http://www.gutenberg.org/files/100/100-0.txt`

For each question, give the answer in a .txt file along with the list of command(s) used. Results should be unique and case-insensitive. For example if you were given a list containing the words: "this" and "THIS" and "This", and were asked to find all the words the began with "th" and ended in "s," the result should contain one word: "This."

Find:

(a) [2 marks] All words that start with a 'th' and end in a vowel

(b) [2 marks] All words that are 14 characters long

(c) [2 marks] All questions asked by Cleopatra

(d) [2 marks] The number of words containing two z's