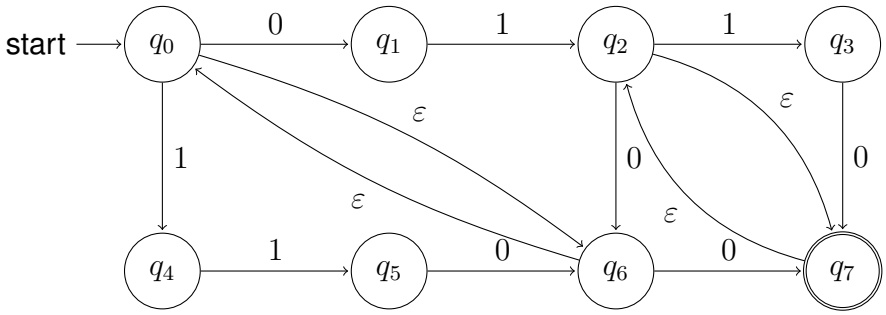
Assignment 2

Due Friday February 17th, 2017

Submission Instructions: Submit solutions in a single PDF via OWL. Assignments are due at 11:55 pm (Eastern Time) on the date listed above. Assignments submitted more than 48 hours late will not be accepted, and a mark of zero (0) will be recorded. See the course outline for details.

1. [5 marks] DFAs, NFAs, and ε-NFAs

(a) [5 marks] The following NFA is an ε-NFA. That is, it contains transitions that occur on no input (i.e., the empty string). Convert this ε-NFA into an equivalent non ε-NFA, i.e., one that does not contain ε-transitions.



2. [9 marks] Regular Expressions

(a) [3 marks] Construct an ε-NFA that represents the language defined by the following regular expression:

$$R = \left((10^*110^+)^*10(11)^+\right) + \left(1101^+\right)$$

Note the '$+$' notation can have two different meanings depending on context. a^+ means "one or more copies of a", while $a + b$ means a *or* b.

(b) [3 marks] Let R_1 and R_2 be the following regular expressions:

$$R_1 = (00 + 11)^+$$

$$R_2 = (11)^*$$

Using set builder notation, describe the following language:

$$L = \{L_1/L_2\},$$

where L_1 and L_2 are the languages represented by regular expressions R_1 and R_2, and where A/B denotes the right quotient of languages A and B.

(c) [3 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, write a sentence or two to describe the language generated by the following regular expression:

$$R = (1 + 0)^* 1^+ + (01 + 001)^*$$

3. [18 marks] Pumping Lemma

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

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

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

$$L_b = \{10^i 10^{i+1} 10^{i+2} 1 ... 0^{i+n} 1 : i, n > 1\}$$

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

$$L_c = \{0^n : n = m^3 \text{ for } m \geq 2 \text{ , i.e., } n = 8, 27, 64...\}$$

4. [8 marks] Context-free Grammars

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

(a) [2 marks]

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

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

(b) [2 marks]

$$L_c = ((01^* + 0^*) + 11)^*$$

(c) [2 marks]

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

[2 marks] Consider the following context-free grammar:

$$S \rightarrow \text{a}S\text{b}|\text{b}Y|Y\text{a}$$
$$Y \rightarrow \text{b}Y|\text{a}Y|\epsilon$$

In one or two sentences, describe the language generated by this grammar.

5. [10 marks] Regex in Practice

For this question you will perform string matching on a large text file. Download the Complete Works of Shakespeare, a text file containing the entire collection of Shakespearean works, located here:

http://www.gutenberg.org/cache/epub/100/pg100.txt

Answer the following questions using regular expressions in Bash command-line utilities such as grep, sed, and awk. A handy way to download the file directly from the command line is

```
wget http://www.gutenberg.org/cache/epub/100/pg100.txt
```

If you are a Windows user you can get access to a Bash command line by running Linux in a virutal machine (e.g., Ubuntu in Virtual Box), or by installing Bash on Ubuntu on Windows.

For each question, give the answer in a .txt file and list of command(s) used. Find:

(a) All words that start with `th` and ends with a consonant

(b) All words that are 12 characters long

(c) All questions asked by Cleopatra.

(d) The longest word whose only vowels are `a`'s

(e) The shortest word with two `r`'s

Tip: Ends of lines may contain blanks or non-printable characters. You can match those with the expression `[[:blank:][:cntrl:]]*`