CS3331 – Assignment 1
due Oct. 12, 2021
2-day no-penalty extension until: Oct. 14, 11:55pm
(SRA's cannot be used to extend further)

1. (60pt) For each of the following languages, prove whether it is regular or not. If it is, then:

   - construct a NDFSM for it
   - convert the NDFSM into a DFSM (Note that you do not have to include trap/dead states)
   - minimize the DFSM
   - convert one of the machines into a regular expression (whichever gives a simpler regular expression)

   Show your work. If you can give directly a DFSM, then you don't have to provide a NDFSM. If you provide directly the minimal DFSM, you still need to argue why it is minimal.

   (a) $L_1 = \{w \in \{a,b\}^* \mid w = uvuv \text{ for some } u, v \in \{a,b\}^*\}$.
   (b) $L_2 = \{w \in \{a,b\}^* \mid w = uvu \text{ for some } u, v \in \{a,b\}^*\}$.
   (c) $L_3 = \{w \in \{a,b\}^* \mid w \text{ has } ababa \text{ as substring}\}$.
   (d) $L_4 = \{w \in \{a,b\}^* \mid w \text{ has } ababa \text{ as a suffix}\}$.
   (e) $L_5 = \{w \in \{a,b\}^* \mid w \text{ does not have } ababa \text{ as substring}\}$.
   (f) $L_6 = \{w \in \{a,b\}^* \mid \#_a(w) - \#_b(w) \equiv 0 \pmod 2\}$. ($\#_a(w)$ is the number of a's in w).
   (g) $L_7 = \{w \in \{a,b\}^* \mid \#_a(w) - \#_b(w) = 0\}$.
   (h) $L_8 = L((a+b)^*(a+\varepsilon)b^*)$

2. (20pt) This question concerns the Multi-Pattern Searching (see slides 43-45 of Regular Expressions chapter) where the automaton reports each occurrence of a pattern in the given set. It does so by going to an accepting state every time the end of a pattern in found in the input text. Assume the English alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \ldots, \mathtt{z}\}$.

   (a) (10pt) Construct the minimal DFSM to solve the multi-pattern searching problem for the patterns $p_1 = \mathtt{int}$, $p_2 = \mathtt{uint}$. (This DFSM can be used for keyword identification in a language that has such keywords.)
   (b) (10pt) Denote the language accepted by the DFSM at (a) by L. Describe the equivalence classes of $\approx_L$.

   Show your work. You can use Thomson's construction, or directly build a correct NDFSM or a correct DFA, or even the minimal DFA. Either way, you need to argue why your construction is correct. (No formal proof is required, only logically sound arguments.)

3. (20pt) Show that the following problem is decidable:

   Given a DFSM M, is it true that the language accepted by M consists of floating-point numbers (written in base 10) such that all floating-point numbers are in $L(M)$ with the possible exception of finitely many ones?

You are allowed to use any of the following:

– closure properties: union, concatenation, Kleene star, complement, intersection, difference

– conversion algorithms between DFSM, NDFSM, regular expressions, and regular grammars (see the last slide of Ch.7: Conversions)

– decision algorithms: membership, emptiness, finiteness, totality, equivalence, minimality.

Explain clearly which of the above closure property and algorithm you have used at each step. Any other construction or algorithm should be described in the assignment.

READ ME! Submit your solution as a single pdf file on `owl.uwo.ca`. Solutions should be typed but high-quality hand-written solutions are acceptable. Make sure you submit everything as a single pdf file.

JFLAP: You are allowed to use JFLAP to help you solve the assignment. You still need to explain clearly your solution. Also, make sure you understand what it does; JFLAP will not be available during exams!

LaTeX: For those interested, the best program for scientific writing is LaTeX. It is far superior to all the other programs, it is free, and you can start using it in minutes; here is an introduction: `https://tobi.oetiker.ch/lshort/lshort.pdf`