CSC 520 Fall 2023
HW #3, Finite Automata (FA) and Regular Languages
25 points/10 extra credit points

Submit a .zip archive containing 3 files:
- A .txt file containing only your honor code affirmation, with your actual name substituted for <*name*>: **On my honor as an SFSU student, I, <*name*>, have neither given or received inappropriate help with this homework assignment.**
- A .jff file containing your solution to problem #1.
- A .txt file containing your solution to problem #2.

1. 12.5 points/8 extra credit points for a minimal correct implementation. Submit your solution as a .jff file.

Σ={C,A,G,T}, L = { *w* : the first and last symbols in *w* are *A's* ; $\#_C(w)+1 = \#_G(w) + \#_T(w) + \#_A(w)$; $\#_C(w) \leq 2$ }.As examples, ACA $\in$ L; *w* = AA $\notin$ L because $\#_C(w)+1 \neq \#_G(w) + \#_T(w) + \#_A(w)$; GCA $\notin$ L because it does not start with A; and *w* = ATCCTCA $\notin$ L because $\#_C(w) > 2$. You may find it helpful to adopt a state naming conventions that reflects the counts of *C's* vs. *A's*, *G's*, and *T's* read from the input if the FA is in that state. For example, the state name "2/1" could be used to mean that if the FA is in that state, the input processed contained 2 *A's*, *G's*, or *T's*, and 1 C.

Construct a deterministic JFLAP FA that recognizes L. Remember that to be deterministic, an FA must have no ε-transitions (rendered as λ-transitions by default in JFLAP), and not have multiple transitions on the same input symbol.

2. 12.5 points/2 extra credit points for explaining why an FA could or could not accept L defined below. Submit your solution as a .txt file

Σ = {C,A,G,T}, L = { *w* : *w* starts with CA and ends with GT**;** $\#_{CA}(w) \% 2 = (\#_{GT}(w) + 1) \% 2$**;** $\#_C(w) > \#_G(w)$ }. As examples, CACATGT $\in$ L; *w* = CACAGTGT $\notin$ L because $\#_C(w) = \ngeq \#_G(w)$; TCACAGT $\notin$ L because it does not start with TA; *w* = CACATCAGT $\notin$ L because $\#_{CA}(w)$ and $\#_{GT}(w)$ are both odd; and *w* = CACAGTCTGT $\notin$ L because $\#_{CA}(w)$ and $\#_{GT}(w)$ are both even.

Prove that L is not a regular language using the RL pumping lemma.

Start by defining a "long" string *S* such that *S* $\in$ L, |*S*| $\geq$ N. Remember that for |*S*| $\geq$ N, **S must be defined using N, or an expression derived from N, as a repetition operator**.

Then show that the RL pumping lemma guarantee for long strings in infinite regular languages does not hold for S: there is **no substring *R*** contained within the first N symbols of *S*, *R* $\neq \varepsilon$, such that *R* could be removed, or replicated an arbitrary number of times (i.e., pumped), and the resulting string *S'* will also be in L.