CSC 520 Fall 2023
Midterm #3 – Languages and Automata
125 points/10 extra credit points

Submit a .zip archive that contains 4 files:

- A plain text .txt file that affirms the honor code: `On my honor as an SFSU student, I, <name>, have neither given or received inappropriate help with this midterm exam.`
- A plain text .txt file containing your solution to problem 1.
- A .jff file containing your solution to problem 2. (To ensure that you have submitted the correct file, you can also email your .jff to mpico@sfsu.edu.)
- A plain text .txt file containing your solution to either problem 3.a or problem 3.b, but not both.

1. (50 points. Submit your solution as a .txt file) $\Sigma = \{C,A,G,T\}$, $L = \{ w : w = C^{c}A^{a^1}GA^{a^2}GA^{a^3}T^{t}$, c % 2 = 1; t = min($a_1$, a_2, a_3); a_1, a_2, $a_3 > 0\}$. For example, $C^3A^5GA^3GA^7T^3 \in L$; $A^5GA^3GA^7T^3 \notin L$ because it starts with 0 *C's*, and 0 is an even number; and $CA^5GA^2GA^7T^3 \notin L$ because its minimum A group size is 2, but there are 3 *T's* at the end.

Prove that L $\notin$ RLs using the regular language pumping lemma. Start by defining a string *S* such that *S* $\in$ L and |*S*| ≥ N, N ≥ 1. Please do not use any of the example strings above in your proof, or define *x*, *y*, and *z* substrings in the style found in the 520 archives. Note that for *S* to be at least N symbols long, it must use N, and/or an expression derived from N, as a repetition operator. For example, $CA^{a^1}GA^{a^1+1}GA^{j+a^1}T^{a^1}$ could not work as an *S*, both because a^1 is undefined outside the characteristic function for L, and because this *S* is not defined in terms of N — thus it is meaningless to ask if it is at least as long as N.

Other points to bear in mind:

- The only valid assumption about N is that is at least 1. For example, consider the language $L = \{ w : w = G^jT^k, j \% 2 = 1, j < k\}$. $S = G^NT^{N+1}$ would not work, because you cannot assume that N is odd.
- After defining *S*, the next step is to show that for all non-null substrings *R* within the first N symbols of *S*, there is some way to pump (remove or replicate) *R* to generate *S'* $\notin$ L.

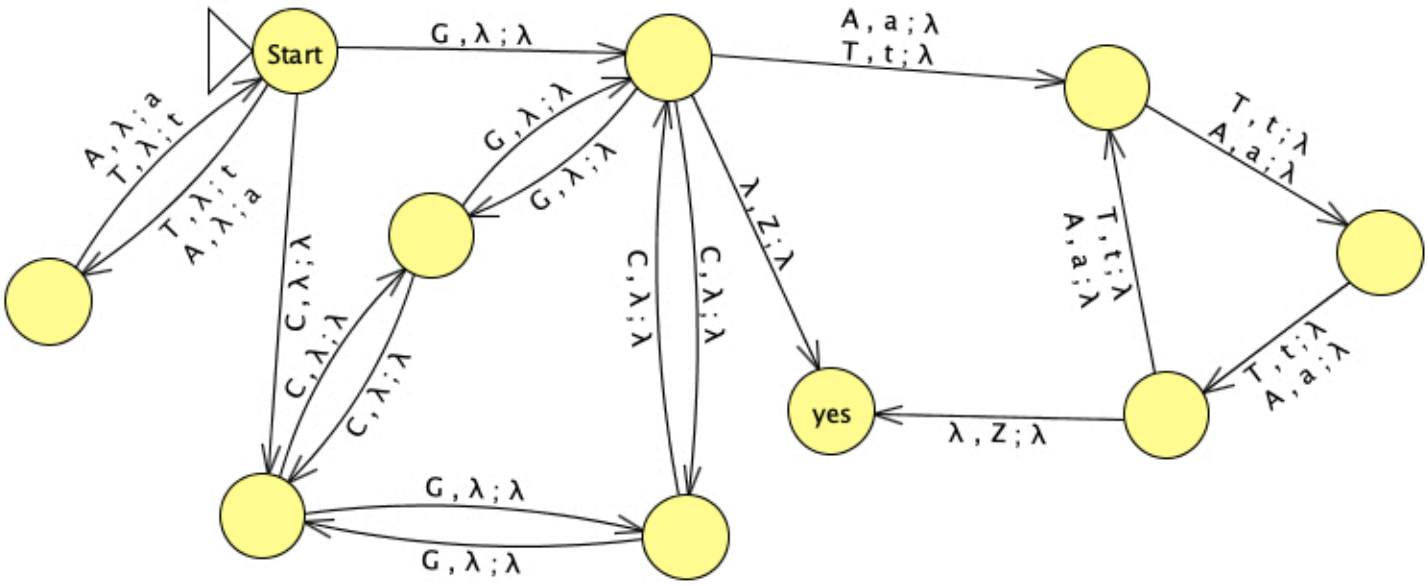
2. (**50 points/10 extra credit points for a minimal correct solution. Submit your solution as a .jff file**) $\Sigma = \{C,A,G,T\}$, $L = \{ w : w$ starts with T or C, and ends with the same symbol that it starts with; $\#_A(w) \leq 1$; if *w* starts with T then $\#_C(w) \leq 1$, and $\#_C(w) + 1 = \#_G(w) + \#_T(w)$; else if *w* starts with C then $\#_C(w) = 2 = \#_G(w) + \#_T(w)\}$. For example, TCAT ∈ L, but TCAG ∉ L because it ends with G, and TCCT ∉ L because $\#_C(TCCT) + 1 \neq \#_G(TCCT) + \#_T(TCCT)$; and CTGC ∈ L, but CTCG ∉ L because it does not end with C.

You may find M3-test_cases.txt useful (also included with the test materials). All the test cases above TCCT should be accepted; the ones below and including TCCT should be rejected. To load the test cases into JFLAP, select Multiple Run from the Input menu, and then select Load Inputs from the buttons on the bottom right.

Remember that a DFA (a deterministic FA) does not use ε-transitions (rendered as λ-transitions in JFLAP), or multiple transitions on the same input symbol. You will lose points if your FA uses nondeterministic transitions.

3. (**25 points. Submit your answer to either 3.a or 3.b, but not both, as a plain text .txt file.**)

3.a **Find a characteristic function for the language decided by Mystery-PDA.jff**, included with the test materials, and shown below. It is not enough to describe how a characteristic function works, you must derive an expression that defines strings in the language accepted by this machine. Note that this PDA is designed to accept by empty stack.



3.b **Conceptual question on transformer/attention Large Language Models (LLMs)**. Imagine that input #1 to an encoder was "The animal did not cross the street because it was too busy."

Give a high-level description the data input to the first layer of the encoder, and how it would differ from the data input to the second layer.

In the first layer, would you expect the attention score for "it" and "animal" to be higher or lower than the score for "it" and "street"?

Suppose input #2 was "The animal did not cross the street because it was too tired ~~busy~~." Would you expect the attention score for "it" and "street" to be higher in the second case than in the first?