CS 100 Lab 11: Formal Languages

Team member(s): ____
(teams of up to two students — only one submission per team)

1. Consider the following context-free grammar for English sentences:

   S → NP VP
   NP → Det Adj N
   VP → Adv V NP
   Det → a | the
   Adj → quick | brown | sad
   N → cow | fox | owl
   Adv → happily | greedily | slowly
   V → jumps | eats | catches

   Given the start symbol S, write down 3 separate sentences that can be generated by this grammar.

2. The grammar from the previous exercise generates a tiny subset of valid English sentences. Drawing from the same vocabulary, and using all the parts-of-speech non-terminals (`Det`, `Adj`, `N`, `Adv`, `V`) from exercise 1, come up with two separate grammars that generate different, valid-English sentence structures. You may add additional non-terminals and terminals (e.g., for punctuation and other parts-of-speech) to your grammars.

   For each grammar you devise, provide one sample sentence.

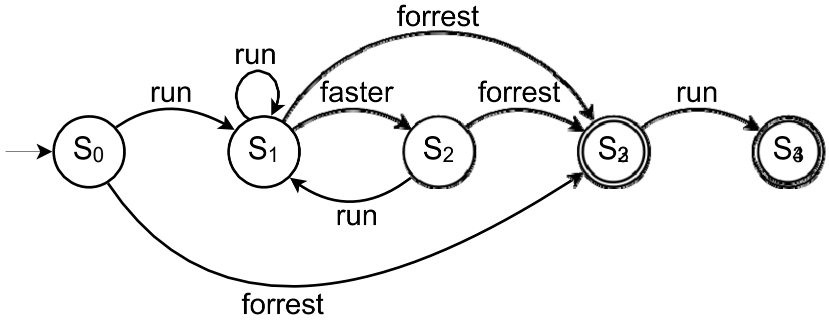
Grammar 1:

Sample sentence:

Grammar 2:

Sample sentence:

3. Consider the following finite-state automaton with start state S_0 and final states S_3 and S_4:



For each of the following sentences, circle the corresponding Y if it is accepted by the FSA, and N otherwise:

run forrest run	Y / N
faster run faster forrest	Y / N
run run run run run	Y / N
run run run forrest	Y / N
run faster run faster	Y / N
run faster forrest faster	Y / N
run faster faster faster forrest	Y / N
run faster run forrest	Y / N
forrest run	Y / N
forrest run faster	Y / N

Write down the regular grammar that corresponds to the FSA:

4. As you may recall from biology, DNA sequences are often represented as strings of the characters A, T, C, and G, which stand for the nucleotide bases Adenine, Thymine, Cytosine, and Guanine used to encode genetic information. A genetic code is made up of triplets of bases (e.g., AGA, CCG, GGT) known as codons, and long sequences known as open reading frames (ORFs) in DNA can be identified by known opening and closing codons.

   For this exercise, you are to draw an FSA that recognizes (accepts) DNA ORFs bounded by the opening codon ATG and the closing codon TAG. Any number and type of codons can occur between them (except for the closing codon), keeping in mind that each codon must be made up of three bases.

   E.g., the following are valid ORFs (spaces inserted for legibility):

   - ATG TAG
   - ATG AGA GAT TCC CAG TGA TAG
   - ATG ATG ATG TAG
   - ATG TGA TGG TAA TAG

   E.g., the following are invalid ORFs (spaces inserted for legibility):

   - ATG TAG TAG
   - GTG TAG
   - GTG ATG TAG
   - ATG TAG TTT TAG
   - ATG TT TAG