CS 332: *Theory of Computation* Prof. Mark Bun
Boston University September 22, 2021

Homework 3 – Due Tuesday, September 28, 2021 at 11:59 PM

Reminder Collaboration is permitted, but you must write the solutions *by yourself without assistance*, and be ready to explain them orally to the course staff if asked. You must also identify your collaborators and write "Collaborators: none" if you worked by yourself. Getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.

Problems There are 5 required problems and 1 bonus problem. Problem 1 will be autograded by AutomataTutor.

1. (**Regular expressions vs. finite automata**) Please log on to AutomataTutor to submit solutions for this question.

   (a) (**Description to regex**) Give regular expressions generating the following languages:

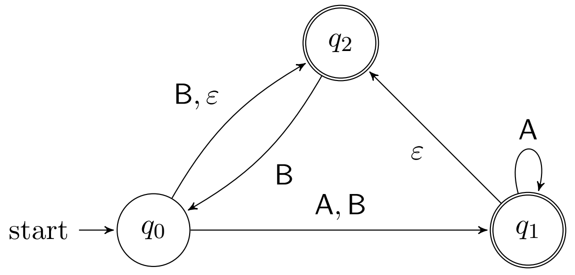
      i. $\{w \in \{0,1\}^* \mid w$ has exactly one 0 and at least two 1's$\}$.

      ii. $\{w \in \{0,1\}^* \mid w$ is any string except for 00 or 000$\}$.

      iii. $\{w \in \{0,1\}^* \mid w$ starts with 0 and has even length or w starts with 1 and has odd length$\}$.

   (b) (**Regex to NFA**) Use the procedure described in class (also in Sipser, Lemma 1.55) to convert $((\mathsf{A} \cup \mathsf{C})^*(\mathsf{T} \cup \mathsf{G})^*(\mathsf{TA} \cup \mathsf{TG}))^*$ to an equivalent NFA. Simplify your NFA.

   (c) (**NFA to regex**) Convert the following NFA (over alphabet $\{\mathsf{A}, \mathsf{B}\}$) to an equivalent regular expression.



2. (**The fault in our stars**)

   (a) If A and B are *finite* languages, what is the maximum cardinality of $A \cup B$? What is the maximum cardinality of $A \circ B$? Explain your answers.

   (b) A *star-free regular expression* is a regex using only the symbols $\emptyset, \varepsilon, \cup, \circ$, and alphabet symbols in R. The *size* of a star-free regex is the number of such symbols it contains. (That is, do not count parentheses.) For a natural number k, let $S(k)$ denote the maximum number of elements in the language generated by a star-free regex R of size k. Prove by induction on k that $S(k) \leq 2^{2^k}$.

   (c) Conclude that a star-free regex always generates a finite language.

3. (**Regex complement**)

   Explain how you could design an algorithm that, given a regex R, constructs a regex S such that $L(S) = \overline{L(R)}$. You may assume you have access to functions such as RegexToNFA, NFAToRegex, NFAtoDFA, etc. implementing the constructions we've seen in class. For example, NFAtoDFA takes as input an NFA and uses the subset construction to return an equivalent DFA. You can assume these functions handle low-level details like parsing parentheses for you. You can also use functions performing the other constructions we've seen in class; just either give a precise reference to the slides or the textbook, or a 1-2 sentence description of the construction to make it completely clear what you are referring to.

4. (**Distinguishing set method**) Let $\Sigma = \{a, b\}$. For each $k \geq 1$, let C_k be the language consisting of all strings that contain an a exactly k places from the right-hand end. That is, $C_k = \{wax_1 \dots x_{k-1} \mid w \in \Sigma^*, x_1, \dots, x_{k-1} \in \Sigma\}$.

   (a) Describe an NFA with $k+1$ states that recognizes C_k in terms of both a state diagram and a formal description.

   (b) Prove that C_k has a pairwise distinguishable set of size 2^k.

   (c) Use part (b) to conclude that no DFA with fewer than 2^k states can recognize C_k.

   (d) Could there be an NFA with fewer than k states recognizing C_k? Explain your answer.

5. (**Non-regular languages**) Prove that the following languages are not regular. You may only use the distinguishing set method and the closure of the class of regular languages under union, intersection, complement, and reverse.

   (a) $L_1 = \{0^n 1^m \mid n, m \geq 0 \text{ and } n = m^2\}$.

   (b) $L_2 = \{0^n 1^m 0^n \mid m, n \geq 0\}$.

   (c) $L_3 = \left\{1^k y \mid y \in \{0,1\}^* \text{ and } |y| = k\right\}$.

   (d) $L_4 = \left\{a^i b^j c^k \mid i, j, k \geq 0 \text{ and if } i = 1 \text{ then } j = k\right\}$.

   (e) Let $\Sigma = \{0, 1, +, =\}$. Let $L_5 = \{x{+}y{=}z \mid x, y, z \text{ are binary nonnegative integers and } x + y = z\}$.

6. (**Bonus problem**) Prove that for every natural number n, there is a language B_n such that a) B_n is recognizable by an NFA with $n+1$ states, but b) If $B_n = A_1 \cup \cdots \cup A_k$ for regular languages $A_1, \dots, A_k$, then at least one of the languages A_i requires a DFA with at least $2^{n/k}$ states.