MATH 4931 / 5931 – SPECIAL TOPICS: THEORY OF COMPUTATION AND FORMAL LANGUAGES

HOMEWORK 2

Spring 2023

due Wednesday, February 22, by the beginning of class

This homework assignment was written in LaTeX. You can find the source code on the course website.

You must explain your reasoning for all of your answers. Correct answers with no justification or explanation will not be accepted.

Students enrolled in Math 4931: Complete any 6 out of the 9 problems.
Students enrolled in MSSC 5931: Complete any 8 out of the 9 problems.

1. Find a DFA for the language $\{w \in \{a,b\}^* : w \notin (a^* \cup b^*)\}$. Draw the diagram of the DFA and give the full formal definition.

2. During the proof that the set of recognizable languages is closed under union, we noted that setting the accept states $F = F_1 \times F_2$ forms a DFA for the intersection of two languages. Define $\mathcal{L} = \{w \in \{a,b\}^* : w \text{ has at least three } a\text{'s and at least two } b\text{'s}\}$. Describe $\mathcal{L}$ as the intersection of two simpler languages, find DFAs for each of them, and use the intersection construction to derive a DFA for $\mathcal{L}$. Draw the diagram of the DFA and give the full formal definition.

3. (a) Find a DFA for the language $\{w \in \{0,1\}^* : w \text{ has length at least 3 and its third symbol is a } 0\}$. Draw the diagram and give the full formal definition.

   (b) Define $\mathcal{L}_n = \{w \in \{0,1\}^* : w \text{ has length at least } n \text{ and its } n\text{th symbol is a } 0\}$. Give the formal definition for a DFA for $\mathcal{L}_n$.

4. Let $\Sigma = \{a,b\}$. For $k \geq 1$, let $\mathcal{C}_k$ be the language consisting of all strings that contain an a exactly k places from the right-hand end. Prove that for each k, no DFA can recognize $\mathcal{C}_k$ with fewer than 2^k states.

5. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be recognizable languages. Define the *perfect shuffle* of $\mathcal{L}_1$ and $\mathcal{L}_2$ to be the words you get by interleaving equal length words from each:

$$\{w : w = u_1v_1u_2v_2\cdots u_kv_k, \text{ where } u_1\ldots u_k \in \mathcal{L}_1 \text{ and } v_1\ldots v_k \in \mathcal{L}_2\}.$$

   Prove that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are recognizable, then so is their *perfect shuffle*.

6. In certain programming languages, comments appear between delimiters such as `/#` and `#/`. Let $\mathcal{C}$ be the language of all valid delimited comment strings. A member of $\mathcal{C}$ must begin with `/#` and end with `#/` but have no intervening `#/`. For simplicity, assume the alphabet for $\mathcal{C}$ is $\{a, b, /, \#\}$. So, for example

$$/\#abbaa/ba\#a/\#ba\#/ \in \mathcal{C} \qquad \text{but} \qquad /\#abba\#/abcs\#/ \notin \mathcal{C}.$$

(a) Give a DFA that recognizes $\mathcal{C}$. A diagram will suffice, you do not need to give the formal definition.

(b) Give a regular expression that matches $\mathcal{C}$.

7. Recall from Homework 1 that $\mathcal{L}^R$ is the language of all words from $\mathcal{L}$, but reversed. There is a theorem (that we need NFAs to prove) that if $\mathcal{L}$ is recognizable by a DFA then so is $\mathcal{L}^R$. However, even without NFAs, it's possible to demonstrate that a DFA for $\mathcal{L}^R$ could end up with exponentially more states that a DFA for $\mathcal{L}$. Give a convincing example of this. (To be the most convincing, you should come up with a family of languages $\mathcal{L}_k$ such that the number of states of $\mathcal{L}_1^R$, $\mathcal{L}_2^R$, $\mathcal{L}_3^R$, etc grows exponentially fast relative to the number of states of $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \ldots$.)

8. For any language $\mathcal{L}$, define

$$\texttt{Prefix}(\mathcal{L}) = \{u : uv \in \mathcal{L} \text{ for some } v \in \Sigma^*\}.$$

Prove that if $\mathcal{L}$ is recognizable then so is $\texttt{Prefix}(\mathcal{L})$.

9. For any word $w = w_1 w_2 \cdots w_n$, define $\texttt{Double}(w) = w_1 w_1 w_2 w_2 \cdots w_n w_n$. For a language $\mathcal{L}$, define $\texttt{Double}(\mathcal{L}) = \{\texttt{Double}(w) : w \in \mathcal{L}\}$. Assume that $\mathcal{L}$ is recognizable, and find a DFA for $\texttt{Double}(\mathcal{L})$. Make sure to give a full formal definition.