University of Toronto Scarborough

CSC B36 | **Term Test 2** | **18 November 2022**

Student Number: : : : : : : : : : : :

Last (Family) Name: ____________________

First (Given) Name: ____________________

Do not turn this page until you are told to do so.

In the meantime, complete the above and read the rest of this cover page.

Aids allowed: None.

Duration: 90 minutes.

There are 6 pages in this test. Each is numbered at the bottom. *When you receive the signal to start, please check that you have all the pages.*

You will be graded on your mastery of course material as taught in class. So you need to demonstrate this. Unless otherwise stated, you must explain or justify every answer.

Answer each question in the space provided. The last page is intentionally left blank in case you need more space for one of your answers. *You must clearly indicate where your answer is and what part should be marked.* **What you write on backs of pages will <u>not</u> be graded.**

Question	Your mark	Out of
1.		3
2.		6
3.		6
4.		3
5.		6
6.		6
Total		30

1. *There is nothing to do!* Your mark for this question will be $\lfloor \max(M_2, M_3) \rfloor / 2$, where M_2 and M_3 are your marks for questions 2 and 3 respectively.

2. Let $\Sigma = \{0, 1\}$. For arbitrary $x, y \in \Sigma^*$, we define $\#_y(x)$ to be $|\{(u, v) : u, v \in \Sigma^* \text{ and } x = uyv\}|$.

   Let $L_2 = \{x \in \Sigma^* :$ 0100 appears exactly once as a substring in x and
   x has no other substring y such that $|y| = 4$ and $|\#_1(y) - \#_0(y)| > 1\}$.

   Give a regex R such that $\mathcal{L}(R) = L_2$. Briefly explain why your regex is correct.

   For full credit, your regex must be as short as possible (not counting implied parentheses).

   Answer:

   $R = (1 + \epsilon)\ (01)^*\ 0100\ (1100)^*\ (110 + 11 + 1 + \epsilon)$

   Explanation:

   The 0100 in the middle matches (the one occurrence of) 0100.

   The parts to the left of 0100 matches the only possible strings that can precede 0100 without violating the second property in the definition of L_2.

   The parts to the righrt of 0100 matches the only possible strings that can follow 0100 without violating the second property in the definition of L_2.

   This regex is 33 characters long (not counting implied parentheses).
   Anyone who gives a shorter correct answer deserves a bonus!

3. Let $\Sigma = \{0, 1\}$. For arbitrary $x, y \in \Sigma^*$, we define $\#_y(x)$ to be $|\{(u, v) : u, v \in \Sigma^* \text{ and } x = uyv\}|$.

   Let $L_3 = \{x \in \Sigma^* :$ 111 appears exactly once as a substring in x and x has no other substring y such that $|y| = 3$ and $|\#_1(y) - \#_0(y)| > 1\}$.

   Using as few states as possible, give a DFSA M such that $\mathcal{L}(M) = L_3$.
   Justify the correctness of your DFSA with a state invariant.

   Answer: Here's our state invariant.

$$\delta^*(q_0, x) = \begin{cases} q_0 & \text{iff } x = \epsilon; \\ q_1 & \text{iff } P(x),\ Q(x) \text{ and } x \text{ ends with 1 but not 11;} \\ q_2 & \text{iff } P(x),\ Q(x) \text{ and } x \text{ ends with 11;} \\ q_3 & \text{iff } P(x),\ Q(x) \text{ and } x \text{ ends with 0 but not 00;} \\ q_4 & \text{iff } P(x),\ Q(x) \text{ and } x \text{ ends with 00;} \\ q_5 & \text{iff } x \in L_3 \text{ and } x \text{ ends with 11;} \\ q_6 & \text{iff } x \in L_3 \text{ and } x \text{ ends with 1 but not 11;} \\ q_7 & \text{iff } x \in L_3 \text{ and } x \text{ ends with 00;} \\ q_8 & \text{iff } x \in L_3 \text{ and } x \text{ ends with 0 but not 00,} \end{cases}$$

   where the predicates P and Q (on Σ^*) are defined by:

   $P(x)$: 111 is **not** a substring of x.

   $Q(x)$: x has no substring y such that $|y| = 3$ and $|\#_1(y) - \#_0(y)| > 1$.

   You should be able to draw the DFSA from the above state invariant.

4. *There is nothing to do!* Your mark for this question will be $\lfloor \max(M_5, M_6) \rfloor / 2$, where M_5 and M_6 are your marks for questions 5 and 6 respectively.

5. For this question, all variables take on values in $\mathbb{N}$. Let $L_5 = \{0^{2p}1^q0^{2r}1^s : p+q = r+s \text{ and } p \geq s\}$.

   (a) **[2 marks]** Explain why $L_5 = \{0^{2i}0^{2j}1^k0^{2k}0^{2j}1^i\}$.

   Answer:

   For $p+q = r+s$ and $p \geq s$, let $i = s$, $j = p - s$ $(= r - q)$ and $k = q$.

   Then $i, j, k \in \mathbb{N}$.

   Solving for p, q, r, s in terms of i, j, k, we get: $p = i + j$, $q = k$, $r = k + j$, $s = i$.

   So,
   $$\begin{aligned} & 0^{2p}1^q0^{2r}1^s \\ = & \ 0^{2(i+j)}1^k0^{2(k+j)}1^i \\ = & \ 0^{2i}0^{2j}1^k0^{2k}0^{2j}1^i \end{aligned}$$
   as wanted.

   (b) **[4 marks]** Give a CFG G such that $\mathcal{L}(G) = L_5$. No justification is required.

   For full credit, your CFG must have as few productions as possible.

   Hint: Use the result of part (a), even if you did not complete that part.

   Answer:

   Here's our CFG.

   $S \to 00S1,\ A$

   $A \to 00A00,\ B$

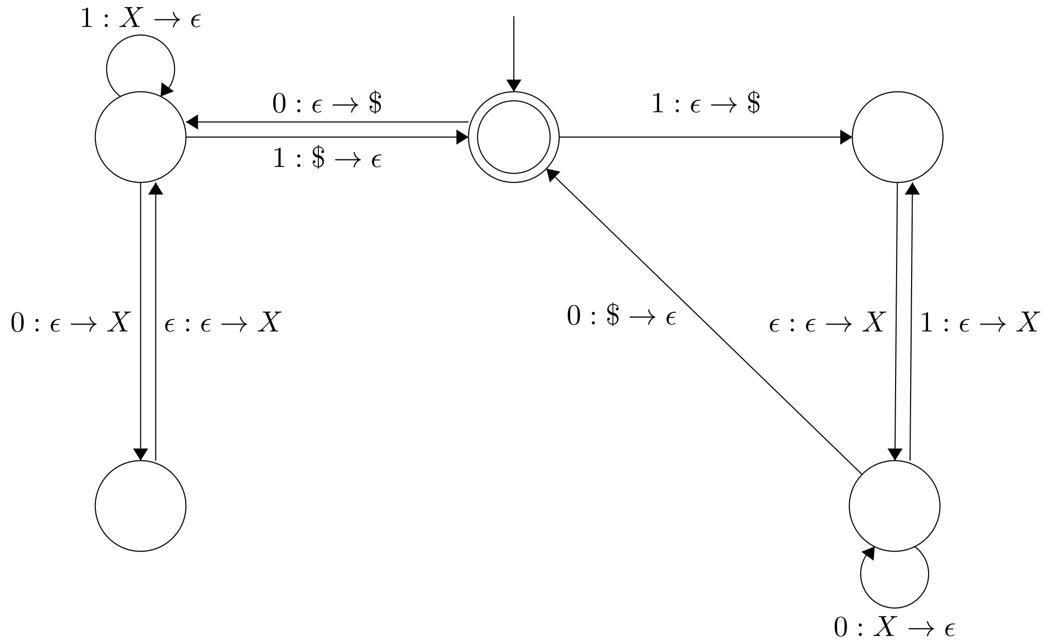
   $B \to 1B00,\ \epsilon$

   This CFG has 6 productions.

   Explanation/Design:

   - S generates $L_5 = \{0^{2i}0^{2j}1^k0^{2k}0^{2j}1^i\}$.
   - A generates $\{0^{2j}1^k0^{2k}0^{2j}\}$.
   - B generates $\{1^k0^{2k}\}$.

6. Let $\Sigma = \{0, 1\}$. Consider the following PDA with input alphabet Σ.



Describe the language accepted by this PDA. Do this without explicit reference to the PDA. I.e., do not answer "the language is $\mathcal{L}(M)$, where M is the above PDA".

Hint: Your answer should include the phrase "nonempty proper prefix".

Answer: The language accepted by the PDA is $(L_0 \cup L_1)^*$, where

$$L_0 = \{x \in \Sigma^* : \; 2 \cdot \#_0(x) = \#_1(x) + 1, \text{ and for every nonempty proper prefix } y \text{ of } x,\ 2 \cdot \#_0(y) > \#_1(y) + 1\},$$

$$L_1 = \{x \in \Sigma^* : \; 2 \cdot \#_1(x) = \#_0(x), \text{ and for every nonempty proper prefix } y \text{ of } x,\ 2 \cdot \#_1(y) > \#_0(y)\}.$$

Note that every nonempty string in L_0 must start with 0 and end with 1, and every nonempty string in L_1 must start with 1 and end with 0.

Brief explanation:

- L_0 is the set of nonempty strings that are accepted by the left side of the PDA.
- L_1 is the set of strings that are accepted by the right side of the PDA.

This page is intentionally left blank in case you need more space for one of your answers.