CS 208 : Automata Theory and Logic Spring 2024

Homework 3

Due: 24^{th} Mar, 2024 *Max Marks: 60*

Instructions:

- *Please start writing your* ***solution to each homework problem on a fresh page.***
- *For each homework problem, you must scan your solution and upload a separate PDF file on Moodle. Please check Moodle for detailed instructions on file naming and uploading instructions.*
- *Be brief, complete, and stick to what has been asked.*
- *Untidy presentation of answers, and random ramblings will be penalized by negative marks.*
- *Unless asked for explicitly, you may cite results/proofs covered in class without reproducing them.*
- ***If you need to make any assumptions, state them clearly.***
- ***Do not copy solutions from others. All detected cases of copying will be reported to DADAC with names and roll nos. of all involved. The stakes are high if you get reported to DADAC, so you are strongly advised not to risk this.***

---

1. Many roads to non-regularlity — 20 points

In this question, we will look at two different ways of proving languages as non-regular. Consider the following two languages over $\Sigma = \{0, 1\}$:

$L_1 = \{0^i 1^j \mid i, j \in \mathbb{N},\ i + j^2 \text{ is a prime number}\}$. For example, $011, 00111 \in L_1$, but $0011, 0111 \notin L_1$

$L_2 = \{w \mid w \text{ differs from } w^R \text{ in exactly two positions}\}$, where w^R is the reverse of the word w. For example, $1110, 1010111 \in L_2$, but $1010, 10100101 \notin L_2$.

1. *[5 + 5 marks]* Show that L_1 and L_2 are not regular by application of the Pumping Lemma (the version mentioned in Tutorial 4, question 2).
2. *[5 + 5 marks]* Show that L_1 and L_2 are not regular by application of the Myhill-Nerode theorem. In each case, you must provide an infinite set of words (suitably described) such that for every pair of words w_1, w_2 in your set, $w_1 \not\sim_{L_i} w_2$, where $\sim_{L_i}$ is the Nerode equivalence for L_i. You must also indicate a distinguishing word x for each w_1, w_2 above such that $w_1 \cdot x \in L_i$ and $w_2 \cdot x \notin L_i$, or vice versa. Answers without the distinguishing words will fetch no marks.

2. Grammatically speaking ... — 20 points

[Please solve this problem after attending Monday's (Mar 18) class.]

Consider the context-free grammar G shown below, where $\{S, A, B\}$ is the set of non-terminals, S is the start symbol and $\{0, 1\}$ is the set of terminals.

$$
\begin{array}{rcl}
S & \rightarrow & AB \mid SS \mid 1S \mid 1 \\
A & \rightarrow & 0S \mid 1B1 \mid \varepsilon \\
B & \rightarrow & 1S \mid 1B
\end{array}
$$

Let $L(G)$ denote the language generated by G.

1. *[2.5+2.5 marks]* Give words $w_1, w_2 \in \{0,1\}^*$ such that (a) each of w_1 and w_2 contains at least one occurrence of 0 and one occurrence of 1, (b) $w_1 \in L(G)$, and (c) $w_2 \notin L(G)$. For w_1, you must give a derivation tree to support your claim that $w_1 \in L(G)$. For w_2, you must give justification why $w_2 \notin L(G)$. Simply giving w_1 and w_2 without justification will fetch no marks.
2. *[5 marks]* Construct another context-free grammar G' that uses only a single non-terminal symbol S and terminal symbols $\{0,1\}$ such that $L(G) = L(G')$. You must provide justification why you think $L(G) \subseteq L(G')$ and $L(G') \subseteq L(G)$.
   [Hint: Think of the language generated by each of the non-terminal symbols in the given grammar. Try to see if you can replace some of these non-terminals in the given production rules by something else that refers to only the start symbol and terminal symbols.]
3. *[10 marks]* We wish to construct a PDA with one state and one symbol in the stack alphabet that accepts $L(G)$ *by the empty stack.* If you think this is possible, list the transitions of the PDA assuming the unique state is q and the unique stack symbol is X (also the initial symbol on stack). In this case, you must also give justification why your PDA, say P, accepts the same language as $L(G)$, i.e. $N(P) \subseteq L(G)$ and $L(G) \subseteq N(P)$. If you think this is not possible, give justification in support of your answer. Feel free to use non-deterministic PDAs for this question.
   [Hint: Think of the grammar with one non-terminal symbol that you constructed in part (2) of the question. Think also about the relation between count of 0*s and count of* 1*s in words in $L(G)$.]*

3. To be or not to be regular — 20 points

Let $\Sigma = \{0,1\}$. Consider the language $L = \{w \in \Sigma^* \mid \exists u, v \in \Sigma^*,\ w = u \cdot v \text{ and } n_0(u) = n_1(v)\}$, where $n_i(x)$ denotes the count of times the letter i appears in the word x. For example, $001 \in L$ because we can write $001 = 0 \cdot 01$ and $n_0(0) = 1 = n_1(01)$.

1. *[10 marks]* Is L regular? If so, give a DFA for L along with justification why the DFA accepts L. Otherwise, prove using either the Pumping Lemma or Myhill-Nerode theorem that L is not regular. Answers without justification will fetch no marks.
2. *[10 marks]* In the example shown above, i.e. $w = 001, u = 0, v = 01$, this is the only way to split w such that the condition for inclusion in the language is satisfied. Is it always true that for every word $w \in L$, there is a unique way to split w as $u \cdot v$ such that $n_0(u) = n_1(v)$? If your answer is "Yes", you must give a proof of why this is so. If your answer is "No", you must provide an example of a word $w \in L$ along with two distinct splits $w = u_1 \cdot v_1 = u_2 \cdot v_2$ such that (a) $u_1 \neq u_2, v_1 \neq v_2$, (b) $n_0(u_1) = n_1(v_1)$, and (c) $n_0(u_2) = n_1(v_2)$. Simply answering "Yes" or "No" for this question will fetch no marks.