CS 3240: Languages and Computation

Problem Set 3 Solutions

Problem 1

Show that the following languages are *not* regular.

- $L_1 = \{www : w \in \{a,b\}^*\}$.

  Solution: We prove that L_1 is nonregular by the pumping lemma. Let p be any given positive integer, and let $s = 0^p10^p10^p1$. Clearly $s \in L_1$ (with $w = 0^p1$), and $|s| = 3p+3 > p$. We show that no matter how s is decomposed into three parts, at least one of the three conditions in the pumping lemma fails to hold, thus the language is nonregular.

  Write $s = xyz$, where $y \neq \varepsilon$ and $|xy| \leq p$. Since s starts with p 0's and $|xy| \leq p$, x and y can consists of only 0's; and since $y \neq \varepsilon$, y consists of at least one 0. Thus $xy^2z = 0^q10^p10^p1$ where $q > p$, and hence $xy^2z \notin L_1$, contradicting the pumping lemma. Therefore L_1 is nonregular.

  □

- The languages in Parts (a), (c) and (d) of Problem 1.46 on Page 90 of Sipser.

  - $L_2 = \{0^n1^m0^n : m, n \geq 0\}$.

    Solution: We prove that L_2 is nonregular by the pumping lemma. Let p be any given positive integer, and let $s = 0^p10^p$. Clearly $s \in L_2$ (with $n = p$ and $m = 1$), and $|s| = 2p+1 > p$. Write $s = xyz$, where $y \neq \varepsilon$ and $|xy| \leq p$. Since s starts with p 0's and $|xy| \leq p$, just as above y consists of only 0's and at least one 0. Thus $xy^2z = 0^q10^p$ where $q > p$, and hence $xy^2z \notin L_2$, contradicting the pumping lemma. Therefore L_2 is nonregular.

    □

  - $L_3 = \{w \in \{0,1\}^* : w \text{ is } \textit{not} \text{ a palindrome}\}$.

    Solution: If L_3 were regular, then by the closure under complementation its complement $\overline{L_3}$ would also be regular. We show that $\overline{L_3}$ is nonregular, from which it follows that L_3 is nonregular.

    We use the pumping lemma to show that $\overline{L_3}$, the language consisting of all palindromes over $\{0,1\}^*$, is nonregular. Let p be any given positive integer, and let $s = 0^p110^p$. Clearly $s \in \overline{L_3}$ (as s is a palindrome), and $|s| = 2p+2 > p$. Write $s = xyz$, where $y \neq \varepsilon$ and $|xy| \leq p$. Since s starts with p 0's and $|xy| \leq p$, just as above y consists of only 0's and at least one 0. Thus $xy^2z = 0^q110^p$ where $q > p$, and hence $xy^2z \notin \overline{L_3}$ (as it is not a palindrome), contradicting the pumping lemma. Therefore $\overline{L_3}$ is nonregular.

    □

  - $L_4 = \left\{wtw : w, t \in \{0,1\}^+\right\}$.

    Solution: We prove that L_4 is nonregular by the pumping lemma. Let p be any given positive integer, and let $s = 0^p110^p1$. Clearly $s \in L_4$ (with $w = 0^p1$ and $t = 1$), and $|s| = 2p+3 > p$. Write $s = xyz$, where $y \neq \varepsilon$ and $|xy| \leq p$. Since s starts with p 0's and $|xy| \leq p$, just as above y consists of only 0's and at least one 0. Thus

$xy^2z = 0^q110^p1$ where $q > p$. We claim that xy^2z can not be written in the form wtw, where $w, t \in \{0,1\}^+$. Write $x^2z = utv$, where u, t, v are all nonempty. If the middle substring t contains the first 1 in xy^2z, then u would end with 0 and v would end with 1, thus $u \neq v$. If t does not contain the first 1 in xy^2z, then $|u| \geq q+1$ and $|v| \leq p+1$, and since $q > p$, $|u| > |v|$ and hence $u \neq u$. Therefore $xy^2z \notin L_4$, and thus L_4 is nonregular.

□

Problem 2

Problem 1.31 on Page 88 of Sipser. Show that regular languages are closed under reversal, that is, if A is regular then so is A^R.

Solution: Let A be any regular language, and let M be a DFA recognizing A. To show that A^R is regular, we construct an NFA N that recognizes A^R, which we describe as follows.

In N we keep all the states of M and *reverse the direction* of all the transition arrows in N. We set the accept state of N to be the start state of M. We also introduce a new state q_0 as the start state of N which goes to every accept state of M by an ε-transition.

Claim 1 *$A^R \subseteq L(N)$.*

Proof: For every $w \in A = L(M)$, the path that w follows in M starts at the start state of M and ends at one of M's accept states. If we input w^R to N, it starts at the new start state q_0 and forks into different paths starting at each of the accept state of M. The computation can reach the accept state of N (that is, the start state of M) by following the same path as w in M but in the opposite direction. Thus $w^R \in L(N)$. □

Claim 2 *$L(N) \subseteq A^R$.*

Proof: For every $x \in L(N)$, there is a path P starting at q_0, going through some accept state of M, say q_a, and stopping at the unique accept state of N (that is, the start state of M). If we input x^R to M, we can go from M's start state to q_a in M by following the path P in the opposite direction. Thus $x \in A^R$. □

Therefore, $L(N) = A^R$ and A^R is regular.

□

Problem 3

Problem 1.48 on Page 90 of Sipser.

Solution: We observe that a string w over $\{0,1\}^*$ contains an even number of occurrences of 01 and 10 if and only if w starts and ends with the same symbol. (Prove this.) Therefore the language D is described by the regular expression $0\Sigma^*0 \cup 1\Sigma^*1$, and hence regular.

□

Problem 4

Give context-free grammars that generate the following languages.

- The languages in Parts (b) and (c) of Exercise 2.4 on Page 128 of Sipser.

  - $\{w : w$ starts and ends with the same symbol$\}$.
    Solution:

$$\begin{aligned} S &\rightarrow 0R0 \mid 1R1 \mid \varepsilon \\ R &\rightarrow 0R \mid 1R \mid \varepsilon \end{aligned}$$

$\square$

  - $\{w :$ the length of w is odd$\}$.
    Solution:

$$S \rightarrow 0 \mid 1 \mid 00S \mid 01S \mid 10S \mid 11S$$

$\square$

- $\left\{ww^R : w \in \{0,1\}^*\right\}$.
  Solution:

$$S \rightarrow 0S0 \mid 1S1 \mid \varepsilon$$

$\square$

- The language in Problem 2.9 on Page 129 of Sipser.
  $\left\{a^i b^j c^k : i = j \text{ or } j = k\right\}$.
  Solution:

$$\begin{aligned} S &\rightarrow E_{ab}C \mid AE_{bc} \\ E_{ab} &\rightarrow aE_{ab}b \mid \varepsilon \\ E_{bc} &\rightarrow bE_{bc}c \mid \varepsilon \\ A &\rightarrow aA \mid \varepsilon \\ C &\rightarrow cC \mid \varepsilon \end{aligned}$$

$\square$

- The language in Part (c) of Problem 2.28 on Page 129 of Sipser.
  $\{w :$ the number of a's is at least the number of b's$\}$.
  Solution:

$$S \rightarrow SS \mid aSb \mid bSa \mid a \mid \varepsilon$$

$\square$

- The set of strings over $\{a, b\}$ with twice as many a's as b's.
  Solution:

$$S \rightarrow SaSaSbS \mid SaSbSaS \mid SbSaSaS \mid \varepsilon$$

$\square$

Problem 5

Problem 2.19 on Page 130 of Sipser. For extra credit, give a rigorous proof of your claims.

Solution: The given grammar G describes exactly the language consisting of all strings *not* of the form $a^n b^n$. (Prove this by induction.) Thus $\overline{L(G)} = \{a^n b^n : n \geq 0\}$, and is generated by the grammar

$$S \rightarrow aSb \mid \varepsilon.$$

□

Problem 6 (Extra Credit)

Parts (b) and (d) of Problem 2.6 on Page 129 of Sipser.

- The complement of $\{a^n b^n : n \geq 0\}$.

  Solution: The grammar in Problem 5. (I meant to assign a different problem here but mistakenly assigned one.) □

- $\left\{x_1\# \cdots \#x_k : k \geq 1,\ x_i \in \{a,b\}^k, \text{ and } x_i = x_j^R \text{ for some } i \text{ and } j\right\}$.

  Solution:

$$\begin{aligned}
S &\rightarrow UPV \\
P &\rightarrow aPa \mid bPb \mid T \mid \varepsilon \\
T &\rightarrow \#MT \mid \# \\
U &\rightarrow M\#U \mid \varepsilon \\
V &\rightarrow \#MV \mid \varepsilon \\
M &\rightarrow aM \mid bM \mid \varepsilon
\end{aligned}$$

□