Theory of Computation Slides
based on Michael Sipser's Textbook

Moses A. Boudourides[1]

Visiting Associate Professor of Computer Science
Haverford College

[1] Moses.Boudourides@cs.haverford.edu

Section 1.4

Nonregular Languages

February 8, 2022

The Pumping Lemma for Regular Languages

Theorem: The Pumping Lemma for Regular Languages

If L is a regular language, then there is a positive integer n (typically, n is the number of states of the DFA accepting L) such that, if $x \in L$ and $|x| \geq n$, then there exist $u, v, w \in \Sigma^*$ such that $x = uvw$ and:

- $|uv| \leq n$,
- $|v| > 0$ (i.e, $v \neq \varepsilon$), and
- for each integer $m \geq 0, uv^m w \in L$.

Corollary

Let the regular language L be accepted by a DFA with n states. Then L is infinite if and only if there is $x \in L$ such that $n \leq |x| < 2n$.

Theorem

The class of regular languages is closed under complement.

A List of Nonregular Languages, I

I.1 $L = \{a^{i^2} \mid i \in \mathbb{Z}, i \geq 0\}$

I.2 $L = \{a^p \mid p \text{ prime}\}$

I.3 $L = \{a^{i^3+3i^2-2i} \mid i \in \mathbb{Z}, i \geq 0\}$

II.1 $L = \{a^i b^i \mid i \in \mathbb{Z}, i \geq 0\}$

II.2 $L = \{a^i b^{pj+q} \mid i \in \mathbb{Z}, i \geq 0\}$ $(p, q \in \mathbb{Z}, p + q \neq 0)$

II.3 $L = \{a^{pi+q} b^j \mid i \in \mathbb{Z}, i \geq 0\}$ $(p, q \in \mathbb{Z}, p + q \neq 0)$

II.4 $L = \{a^i b^j a^i \mid i, j \in \mathbb{Z}, i, j \geq 0\}$

II.5 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i \neq j\}$

II.6 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i > j\}$

II.7 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i < j\}$

II.8 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i \neq pj + q\}$ $(p, q \in \mathbb{Z}, p + q \neq 0)$

II.9 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, j \neq pi + q\}$ $(p, q \in \mathbb{Z}, p + q \neq 0)$

II.10 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, p_1 j + q_1 \leq i \leq p_2 j + q_2\}$
$(p_k, q_k \in \mathbb{Z}, k = 1, 2)$

II.11 $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, p_1 i + q_1 \leq j \leq p_2 i + q_2\}$
$(p_k, q_k \in \mathbb{Z}, k = 1, 2)$

II.12 $L = \{a^{p_1 i + q_1} b^{p_2 j + q_2} \mid i, j \in \mathbb{Z}, i, j \geq 0\}$ $(p_k, q_k \in \mathbb{Z}, k = 1, 2)$

II.13 $L = \{x \in (a + b)^* \mid x \neq x^R\}$

A List of Nonregular Languages, II

III.1 $L = \{xx \mid x \in (a+b)^*\}$

III.2 $L = \{xyx \mid x, y \in (a+b)^+\}$

III.3 $L = \{xx^R \mid x \in (a+b)^+\}$

III.4 $L = \{xx^R y \mid x, y \in (a+b)^+\}$

III.5 $L = \{x \in (a+b)^* \mid n_a(x) = n_b(x)\}$

III.6 $L = \{x \in (a+b)^* \mid n_a(x) \neq n_b(x)\}$

III.7 $L = \{x \in (a+b)^* \mid n_a(x) = p_1 + q_1 n_b(x)\}$
$(p_1, q_1 \in \mathbb{Z}, p_1 \geq 0, q_1 \neq 1)$

III.8 $L = \{x \in (a+b)^* \mid n_a(x) \neq p_1 + q_1 n_b(x)\}$
$p_1, q_1 \in \mathbb{Z}, p_1 \geq 0, q_1 \neq 1)$

III.9 $L = \{x \in (a+b)^* \mid n_b(x) = p_2 + q_2 n_a(x)\}$
$(p_2, q_2 \in \mathbb{Z}, p_2 \geq 0, q_2 \neq 1)$

III.10 $L = \{x \in (a+b)^* \mid n_b(x) \neq p_2 + q_2 n_a(x)\}$
$p_2, q_2 \in \mathbb{Z}, p_2 \geq 0, q_2 \neq 1)$

III.11 $L = \{x \in (a+b)^* \mid p_1 + q_1 n_a(x) \leq n_b(x) \leq p_2 + q_2 n_a(x)\}$
$(p_k, q_k \in \mathbb{Z}, p_k \geq 0, q_k \neq 1, k = 1, 2)$

III.12 $L = \{x \in (a+b)^* \mid p_1 + q_1 n_b(x) \leq n_a(x) \leq p_2 + q_2 n_b(x)\}$
$(p_k, q_k \in \mathbb{Z}, p_k \geq 0, q_k \neq 1, k = 1, 2)$

III.13 $L = \{x \in (a+b)^* \mid n_a(x) \geq n_b(x)\}$

III.14 $L = \{x \in (a+b)^* \mid n_a(x) \leq n_b(x)\}$

IV.1 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, i = j \text{ or } j \neq k\}$

IV.2 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, i \neq j \text{ or } j \neq k\}$

IV.3 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, k = pi + qj\}$ $(p, q \in \mathbb{Z}, pq \neq 0)$

IV.4 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, k \neq qi + j\}$

IV.5 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, k \neq pi + qj\}$ $(p, q \in \mathbb{Z}, pq \neq 0)$

IV.6 $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{Z}, i, j, k \geq 0, k = |i - j|\}$

Examples of Pumping Lemma, I

Example 1 (I.1)

Use the pumping lemma to prove that the language $L = \{a^{i^2} \mid i \in \mathbb{Z}, i \geq 0\}$ is not regular.

Assume that $L = \{a^{i^2} \mid i \in \mathbb{Z}, i \geq 0\}$, i.e., the language of all strings which are perfect squares of a, is regular. Then, by the PL, there exists integer $n \geq 1$, and, for $x = a^{n^2}$, since $|x| = n^2 > n$, the PL would necessitate that $x = a^{n^2} = uvw$, for $u, v, w \in \Sigma^*$.

Now, since $|uv^2w| = |uvw| + |v|, |uvw| < |uv^2w|$ and $|v| < |uv| \leq n$, as $|v| > 0$, we get $n^2 = |uvw| \leq n^2 + n < (n+1)^2$. In other words, the length of uv^2w lies between two squares of two consecutive integers, which implies that uv^2w cannot be a perfect square and, thus, $uv^2w \notin L$. But this is a contradiction, because PL implies that $uv^m w \in L$, for any integer $m \geq 2$, and, in particular, for $m = 2$. Consequently, L cannot a regular language.

Examples of Pumping Lemma, II

Example 2

Use the pumping lemma to prove that the language $L = \{a^i b^i \mid i \in \mathbb{Z}, i \geq 0\}$ is not regular.

Assume that $L = \{a^i b^i \mid i \in \mathbb{Z}, i \geq 0\}$ is regular. Then, by the PL, there exists integer $n \geq 1$, and, for $x = a^n b^n$, since $|x| = 2n > n$, the PL would necessitate that $x = a^n b^n = uvw$, for $u, v, w \in \Sigma^*$.

Claim: v contains only a's (at least one a).

Proof of claim: If v contained both a's and b's, $v = a^p b^q$, for some integers p, q such that $p \geq 0, q \geq 1$. Then, since w follows u (as a string in L), w should only contain b's, say $w = b^s$, for some integer $s \geq 0$. Moreover, since u precedes v, u should contain only a's, say $u = a^r$, for some integer $r \geq 0$. Alltogether, u, v and w, are $x = a^{r+p} b^{q+s} = a^n b^n$, which implies that $r + p = n$. In addition, $|uv| = r + p + q = n + q \geq n + 1 > n$ and this contradicts the first consequence of the PL that $|uv| \leq n$. Therefore, the claim is shown.

To continue with the proof, we have that $u = a^\alpha, v = a^\beta$ and $w = a^\gamma b^\delta$ and, totally, $\alpha+\beta+\gamma = \delta = n$. On the other side, the third consequence of the PL implies that, for all integer $m \geq 0$, $uv^m w \in L$, which gives that $\alpha + m\beta + \gamma$ $= n_a(uv^m w) = n_b(uv^m w) = \delta$. Therefore, $\alpha + m\beta + \gamma = \delta = \alpha + \beta + \gamma$ or $\beta(m-1) = 0$, where $\beta > 0$ (because $|v| > 0$), which generates the contradiction $m = 1$ (because the previous ought to be true for all integer $m \geq 0$). Consequently, L is not a regular language.

Examples of Pumping Lemma, III

Example 3

Use the pumping lemma to prove that the language $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i > j\}$ is not regular.

Assume that $L = \{a^i b^j \mid i, j \in \mathbb{Z}, i, j \geq 0, i > j\}$ is regular. Then, by the PL, there exists integer $n \geq 1$, and, for $x = a^{n+1}b^n$, since $|x| = 2n+1 > n$, the PL would necessitate that $x = a^{n+1}b^n = uvw$, for $u, v, w \in \Sigma^*$.

Similarly to how the claim in the proof of Example 2 was proved (*please, repeat the proof of such a claim in this example and in any similar problem that you are solving!*), it can be shown that v contains only a's (at least one a). Thus, we have that $u = a^\alpha, v = a^\beta$ and $w = a^\gamma b^\delta$ and, totally, $\alpha + \beta + \gamma = n + 1$ and $\delta = n$. On the other side, the third consequence of the PL implies that, for all integer $m \geq 0$, $uv^m w \in L$, and, hence, for the particular value $m = 0$, $uw \in L$, which implies that $\alpha + \gamma = n_a(uw) > n_b(uw) = \delta$. However, since $\alpha + \gamma = n + 1 - \beta$ and $\delta = n$, the last inequality yields that $n + 1 - \beta > n$, i.e., or $\beta < 1$, which is a contradiction, because β is the length of v and by the PL it is assumed that $|v| = \beta > 1$. Consequently, L is not a regular language.

Examples of Pumping Lemma, IV

Example 4

Use the pumping lemma to prove that the language $L = \{xx \mid x \in (a+b)^*\}$ is not regular.

Assume that $L = \{xx \mid x \in (a+b)^*\}$ is regular. Then, by the PL, there exists integer $n \geq 1$, and, for $x = a^n b a^n b$, since $|x| = 2n + 2 > n$, the PL would necessitate that $x = a^n b a^n b = uvw$, for $u, v, w \in \Sigma^*$.

Similarly to how the claim in the proof of Example 2 was proved (*please, repeat the proof of such a claim in this example and in any similar problem that you are solving!*), it can be shown that v contains only a's (at least one a). Thus, the third consequence of the PL implies that, for all integer $m \geq 0$, $uv^m w \in L$, and, hence, for the particular value $m = 2$, $uvvw \in L$, which implies that $uv = vw$. However, since v contains only a's (at least one a), the only case that it could be $uv = vw$ would be that the suffix of w was a, which is a contradiction of the fact that $uvw = a^n b a^n b$, i.e., the fact that, by construction of x, w ends in b. Consequently, L is not a regular language.

Examples of Pumping Lemma, V

Example 5

Use the pumping lemma to prove that the language $L = \{x \in (a+b)^* \mid n_a(x) = n_b(x)\}$ is not regular.

Assume that $L = \{x \in (a+b)^* \mid n_a(x) = n_b(x)\}$ is regular. Then, by the PL, there exists integer $n \geq 1$, and, for $x = a^n b^n$, since $|x| = 2n > n$, the PL would necessitate that $x = a^n b^n = uvw$, for $u, v, w \in \Sigma^*$.

Similarly to how the claim in the proof of Example 2 was proved (*please, repeat the proof of such a claim in this example and in any similar problem that you are solving!*), it can be shown that v contains only a's (at least one a). Thus, we have that $u = a^\alpha, v = a^\beta$ and $w = a^\gamma b^\delta$ and, totally, $\alpha + \beta + \gamma = \delta = n$. On the other side, the third consequence of the PL implies that, for all integer $m \geq 0$, $uv^m w \in L$, which gives that $\alpha + m\beta + \gamma = n_a(uv^m w) = n_b(uv^m w) = \delta$. Therefore, $\alpha + m\beta + \gamma = \delta = \alpha + \beta + \gamma$ or $\beta(m-1) = 0$, where $\beta > 0$ (because $|v| > 0$), which generates the contradiction $m = 1$ (because the previous ought to be true for all integer $m \geq 0$). Consequently, L is not a regular language.

The Myhill–Nerode Theorem

Definition of Indistinguishable Strings in a Language

Let L a language over Σ and let $x, y \in \Sigma^*$. We say that x and y are **indistinguishable with respect to** L and we write $x \approx_L y$ if, for all $z \in \Sigma^*$, either both xz and $yz \in L$ or neither is. Furthermore, $\approx_L$ can be proved to be an equivalence relation on Σ^*.

The Myhill–Nerode Theorem

A language L is regular if and only if the number of equivalence classes of $\approx_L$ is finite.