Lecture 9 – The Pumping Lemma for Regular Languages

COSE215: Theory of Computation

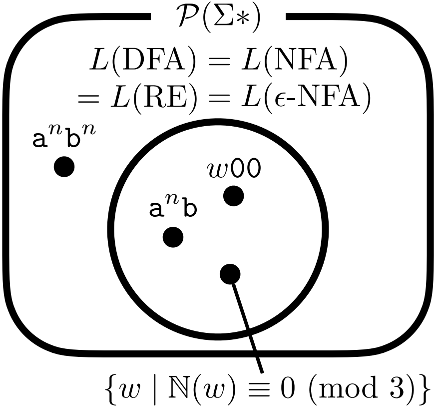
Jihyeok Park

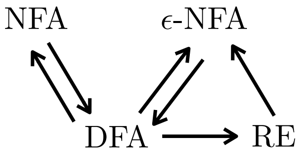
PLRG

2023 Spring

Recall

- Not all languages are regular: e.g., $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$.





- How to prove that a language is **NOT** regular? **Pumping Lemma!**

Contents

Lemma (Pumping Lemma for Regular Languages)

For a given regular language L, **there exists** *a positive integer n such that* **for all** $w \in L$, *if* $|w| \geq n$, **there exists** $w = xyz$ *such that*

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^iz \in L$

$$A \quad = \quad L \text{ is regular}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ 1 \land 2 \land 3$$

Proof of Pumping Lemma

- Let L be a regular language.
- Then, $\exists$ DFA $D = (Q, \Sigma, \delta, q_0, F)$. s.t. $L(D) = L$. Let $n = |Q| > 0$.
- Take any $w = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_m \in L$ s.t. $|w| = m \geq n$.
- Let $p_i = \delta^*(q_0, \mathtt{a}_1 \cdots \mathtt{a}_i)$ for all $0 \leq i \leq m$. Then, $p_0 = q_0 \wedge p_m \in F$.
- Consider the first $n+1$ states: $p_0 \xrightarrow{\mathtt{a}_1} p_1 \xrightarrow{\mathtt{a}_2} \cdots \xrightarrow{\mathtt{a}_n} p_n$.
- By Pigeonhole Principle, there exists $i < j \leq n$ s.t. $p_i = p_j$.
- Split $w = xyz$:

$$\begin{array}{rclcrclcrcl}
x &=& \mathtt{a}_1 \cdots \mathtt{a}_i & \qquad & y &=& \mathtt{a}_{i+1} \cdots \mathtt{a}_j & \qquad & z &=& \mathtt{a}_{j+1} \cdots \mathtt{a}_m \\
|x| &=& i & & |y| &=& j - i > 0 & & |xy| &=& j \leq n \\
\delta^*(q_0, x) &=& p_i & & \delta^*(p_i, y) &=& p_i & & \delta^*(p_i, z) &=& p_m
\end{array}$$

- Then, $\forall i \geq 0.\ \delta^*(q_0, xy^iz) = p_m$ (by induction on i).
- Finally, $\forall i \geq 0.\ xy^iz \in L$.

Proof of Pumping Lemma

- Let L be a regular language.
- Then, $\exists$ DFA $D = (Q, \Sigma, \delta, q_0, F)$. s.t. $L(D) = L$. Let $n = |Q| > 0$.
- Take any $w = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_m \in L$ s.t. $|w| = m \geq n$.
- Let $p_i = \delta^*(q_0, \mathtt{a}_1 \cdots \mathtt{a}_i)$ for all $0 \leq i \leq m$. Then, $p_0 = q_0 \wedge p_m \in F$.
- Consider the first $n + 1$ states: $p_0 \xrightarrow{\mathtt{a}_1} p_1 \xrightarrow{\mathtt{a}_2} \cdots \xrightarrow{\mathtt{a}_n} p_n$.
- By Pigeonhole Principle, there exists $i < j \leq n$ s.t. $p_i = p_j$.
- Split $w = xyz$:

$$\begin{array}{ccc}
x = \mathtt{a}_1 \cdots \mathtt{a}_i & y = \mathtt{a}_{i+1} \cdots \mathtt{a}_j & z = \mathtt{a}_{j+1} \cdots \mathtt{a}_m \\
|x| = i & 1 - |y| = j - i > 0 & |xy| = j \leq n - 2 \\
\delta^*(q_0, x) = p_i & \delta^*(p_i, y) = p_i & \delta^*(p_i, z) = p_m
\end{array}$$

- Then, $\forall i \geq 0.\ \delta^*(q_0, xy^iz) = p_m$ (by induction on i).
- Finally, $\forall i \geq 0.\ xy^iz \in L$ – 3.

Proof of Pumping Lemma

Lemma (Pumping Lemma for Regular Languages)

For a given regular language L, **there exists** *a positive integer n such that* **for all** $w \in L$, *if* $|w| \geq n$, **there exists** $w = xyz$ *such that*

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^iz \in L$

$$A \quad = \quad L \text{ is regular}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \wedge ② \wedge ③$$

Application: Proving Languages are Not Regular

Lemma (Pumping Lemma for Regular Languages)

$$A \;=\; L \textit{ is regular}$$
$$\Downarrow$$
$$B \;=\; \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③$$

$$
\begin{array}{rcll}
A & \Longrightarrow & B & \text{(O)} \\
B & \Longrightarrow & A & \text{(X)} \\
\neg B & \Longrightarrow & \neg A & \text{(O)}
\end{array}
$$

$$
\begin{array}{rcl}
\neg B & = & \forall n > 0.\ \neg(\forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③) \\
 & = & \forall n > 0.\ \exists w \in L.\ \neg(|w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③) \\
 & = & \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \neg(\exists w = xyz.\ ① \land ② \land ③) \\
 & = & \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ \neg(① \land ② \land ③) \\
 & = & \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ \neg(① \land ②) \lor \neg ③ \\
 & = & \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ (① \land ②) \Rightarrow \neg ③
\end{array}
$$

Application: Proving Languages are Not Regular

To prove a language L is **NOT** regular, we need to show that

$$\forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ (①\land ②) \Rightarrow \neg ③$$

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^iz \in L$

Note that $\neg③ = \exists i \geq 0.\ xy^iz \notin L$.

We can prove this by following the steps below:

1. Assume any positive integer n is given.
2. Pick a word $w \in L$.
3. Show that $|w| \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \land$ ② $|xy| \leq n$.
5. $\neg③$ Pick $i \geq 0$, and show that $xy^iz \notin L$ using ① and ②.

Example 1

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $w = \mathtt{a}^n\mathtt{b}^n \in L$.
3. $|w| = n + n = 2n \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \land$ ② $|xy| \leq n$.
5. Then, $\neg$③ $xy^0z \notin L$ because:
   - Since ② $|xy| \leq n$,

     $$x = \mathtt{a}^p \qquad y = \mathtt{a}^q \qquad z = \mathtt{a}^r\mathtt{b}^n$$

     where $0 \leq p, q, r \leq n$ and $p + q + r = n$.
   - Since ① $|y| > 0$, we know $q > 0$.
   - Finally, $xy^0z = xz = \mathtt{a}^p\mathtt{a}^r\mathtt{b}^n = \mathtt{a}^{n-q}\mathtt{b}^n$ ($\because p + q + r = n$).
     But, $\mathtt{a}^{n-q}\mathtt{b}^n \notin L$ ($\because q > 0$). $\square$

Example 2

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

1. Assume any positive integer n is given.
2. Let $w = \mathtt{a}^n\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n \in L$.
3. $|w| = n + n + n + n = 4n \geq n$.
4. Assume any split $w = xyz$ is given, and (1) $|y| > 0 \land$ (2) $|xy| \leq n$.
5. Then, $\neg$(3) $xy^0z \notin L$ because:
   - Since (2) $|xy| \leq n$,

     $$x = \mathtt{a}^p \qquad y = \mathtt{a}^q \qquad z = \mathtt{a}^r\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n$$

     where $0 \leq p, q, r \leq n$ and $p + q + r = n$.
   - Since (1) $|y| > 0$, we know $q > 0$.
   - Finally, $xy^0z = xz = \mathtt{a}^p\mathtt{a}^r\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n = \mathtt{a}^{n-q}\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n$ ($\because p + q + r = n$). But, $\mathtt{a}^{n-q}\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n \notin L$ ($\because q > 0$). $\square$

Example 3

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{\mathtt{a}^l \mathtt{b}^m \mathtt{c}^n \mid l + m \leq n\}$$

1. Assume any positive integer n is given.
2. Let $w = \mathtt{a}^n \mathtt{b}^n \mathtt{c}^{2n} \in L$.
3. $|w| = n + n + 2n = 4n \geq n$.
4. Assume any split $w = xyz$ is given, and (1) $|y| > 0 \land$ (2) $|xy| \leq n$.
5. Then, $\neg$(3) $xy^0z \notin L$ because:
   - Since (2) $|xy| \leq n$,

     $$x = \mathtt{a}^p \qquad y = \mathtt{a}^q \qquad z = \mathtt{a}^r \mathtt{b}^n \mathtt{c}^{2n}$$

     where $0 \leq p, q, r \leq n$ and $p + q + r = n$.
   - Since (1) $|y| > 0$, we know $q > 0$.
   - Finally, $xy^0z = xz = \mathtt{a}^p \mathtt{a}^r \mathtt{b}^n \mathtt{c}^{2n} = \mathtt{a}^{n-q} \mathtt{b}^n \mathtt{c}^{2n}$ ($\because p + q + r = n$). But, $\mathtt{a}^{n-q} \mathtt{b}^n \mathtt{c}^{2n} \notin L$ ($\because q > 0$). $\square$

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{a^{n^2} \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $w = a^{n^2} \in L$.
3. $|w| = n^2 \geq n$.
4. Assume any split $w = xyz$ is given, and 1 $|y| > 0 \wedge$ 2 $|xy| \leq n$.
5. Then, $\neg$3 $xy^2z \notin L$ because:
   - Since 1 $|y| > 0$ and 2 $|xy| \leq n$,

     $$y = a^k$$

     where $1 \leq k \leq n$. Then,

     $$n^2 < n^2 + k \ (\because 1 \leq k) \qquad n^2 + k < (n+1)^2 \ (\because k \leq n)$$

   - Finally, $xy^2z = xyyz = a^{n^2+k} \notin L \ (\because n^2 < n^2 + k < (n+1)^2)$. □

Example 5

Let's prove that L is **NOT** regular:

$$L = \{\mathtt{a}^n \mathtt{b}^k \mathtt{c}^{n+k} \mid n, k \geq 0\}$$

- It is much easier to use **closure properties** under **homomorphisms**.
- Consider a homomorphism $h : \{\mathtt{a}, \mathtt{b}, \mathtt{c}\} \to \{\mathtt{a}, \mathtt{b}\}^*$:

$$h(\mathtt{a}) = \mathtt{a} \qquad h(\mathtt{b}) = \mathtt{a} \qquad h(\mathtt{c}) = \mathtt{b}$$

- Then,

$$h(L) = \{\mathtt{a}^{n+k} \mathtt{b}^{n+k} \mid n, k \geq 0\} = \{\mathtt{a}^n \mathtt{b}^n \mid n \geq 0\}$$

- If L is regular, then $h(L)$ must be regular as well.
- However, we know $h(L)$ is **NOT** regular.
- Therefore, L is **NOT** regular. $\square$

Summary

1. Pumping Lemma for Regular Languages

- Pumping Lemma
- Proof of Pumping Lemma
- Application: Proving Languages are Not Regular
- Examples
  - Example 1: $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$
  - Example 2: $L = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
  - Example 3: $L = \{\mathtt{a}^l\mathtt{b}^m\mathtt{c}^n \mid l + m \leq n\}$
  - Example 4: $L = \{\mathtt{a}^{n^2} \mid n \geq 0\}$
  - Example 5: $L = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^{n+k} \mid n, k \geq 0\}$

Homework #2

- Please see https://github.com/ku-plrg-classroom/docs/tree/main/equiv-re-fa.
- The due date is Apr. 13 (Thu.).
- Please only submit `Implementation.scala` file to **Blackboard**.

Next Lecture

- Equivalence and Minimization of Finite Automata

Jihyeok Park
jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr