COSE215: Theory of Computation

Lecture 2 — Languages and Grammars

Hakjoo Oh
2016 Spring

Alphabet

A finite, non-empty set of symbols, e.g.,

1. $\mathbf{\Sigma = \{0, 1\}}$: the binary alphabet.
2. $\boldsymbol{\Sigma = \{a, b, \ldots, z\}}$: the set of all lowercase letters.
3. The set of all ASCII characters.

String

A finite sequence of symbols chosen from an alphabet, e.g.,

1. $\Sigma = \{0, 1\}$: 0, 1, 00, 01, $\dots$
2. $\Sigma = \{a, b, c\}$: a, b, c, ab, bc, $\dots$

Notations:

- ϵ: the empty string
- wv: the concatenation of w and v
- w^R: the reverse of w
- $|w|$: the length of string w
- $w = vu$: v is a prefix and u a suffix of w.
- Σ^k: the set of strings (over Σ) of length k
- $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \cdots = \bigcup_{k \geq 0} \Sigma^k$
- $\Sigma^+ = \Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \cdots = \bigcup_{k \geq 1} \Sigma^k$

Language

A language L is a set of strings, i.e., $L \subseteq \Sigma^*$ ($L \in 2^{\Sigma^*}$)

When $\Sigma = \{0, 1\}$,

- $L_1 = \{0, 00, 001\}$
- $L_2 = \{0^n 1^n \mid n \geq 0\}$
- $L_3 = \{\epsilon, 01, 10, 0011, 0101, 1001, \ldots\}$
- $L_3 = \{10, 11, 101, 111, 1011, \ldots\}$

Language Operations

- union, intersection, difference: $L_1 \cup L_2, \quad L_1 \cap L_2, \quad L_1 - L_2$
- reverse: $L^R = \{w^R \mid w \in L\}$
- complement: $\overline{L} = \Sigma^* - L$
- concatenation of L_1 and L_2:

$$L_1L_2 = \{xy \mid x \in L_1 \land y \in L_2\}$$

- power:

$$\begin{aligned} L^0 &= \{\epsilon\} \\ L^n &= L^{n-1}L \end{aligned}$$

- closures:

$$L^* = L^0 \cup L^1 \cup L^2 \cup \cdots = \bigcup_{i \geq 0} L^i$$

$$L^+ = L^1 \cup L^2 \cup L^3 \cup \cdots = \bigcup_{i \geq 1} L^i$$

Exercises

1. Consider $L = \{a^n b^n \mid n \geq 0\}$.
    1. $L^2 =$
    2. $L^R =$
2. Prove that $(uv)^R = v^R u^R$ for all $u, v \in \Sigma^+$.

Grammar

Definition

A grammar G is a quadruple $G = (V, T, S, P)$:

- V: a finite set of *variables* (or *non-terminal symbols*)
- T: a finite set of *terminal symbols*
- $S \in V$: the *start* variable
- P: a finite set of productions. A production has the form

$$x \rightarrow y$$

  where $x \in V$ and $y \in (V \cup T)^*$.

Example:

$$G = (\{S\}, \{a, b\}, S, P)$$

$$\begin{array}{lcl} S & \rightarrow & aSb \\ S & \rightarrow & \epsilon \end{array}$$

Applying productions to strings

- $x \to y$: replace x by y, e.g., applying $x \to y$ to the string:

$$w = uxy$$

gives

$$z = uyv$$

In this case, we write $w \Rightarrow z$.

- $w_1 \Rightarrow^* w_n$ iff $w_1 \Rightarrow w_2 \Rightarrow \cdots \Rightarrow w_n$

Example

$$G = (\{S\}, \{a, b\}, S, P)$$

$$\begin{aligned} S &\rightarrow aSb \\ S &\rightarrow \epsilon \end{aligned}$$

- $S \Rightarrow^* aabb$
- $S \Rightarrow^* aaabbb$

A grammar specifies a language

Definition

Let $G = (V, T, S, P)$ be a grammar. Then the set

$$L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$$

is the language generated by G.

- If $w \in L(G)$, then we say the sequence

$$S \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \cdots \Rightarrow w_n \Rightarrow w$$

  a *derivation* of the sentence w.
- $S, w_1, w_2, \ldots, w_n$: *sentential forms*.

Example

$$G = (\{S\}, \{a, b\}, S, P)$$

$$\begin{array}{rcl} S & \rightarrow & aSb \\ S & \rightarrow & \epsilon \end{array}$$

The language of G is

$$L(G) = \{\epsilon, ab, aabb, aaabbb, aaaabbbb, \ldots\} = \{a^n b^n \mid n \geq 0\}.$$