COSE215: Theory of Computation

Lecture 8 — Context-Free Grammars

Hakjoo Oh
2016 Spring

Context-Free Languages

An extension of the regular langauges. Many applications in CS:

- Most programming languages (e.g., C, Java, ML, etc).
- Markup languages (e.g., HTML, XML, etc).
- Essential to design of programming languages and construction of compilers.

Example: Palindromes

- A string is a palindrome if it reads the same forward and backward.
- $L = \{w \in \{0,1\}^* \mid w = w^R\}$
- L is not regular, but context-free.
- Every context-free language is defined by a recursive definition.
  - Basis: ϵ, 0, and 1 are palindromes.
  - Induction: If w is a palindrome, so are $0w0$ and $1w1$.
- The recursive definition is expressed by a context-free grammar.

$$\begin{array}{lcl} P & \rightarrow & \epsilon \\ P & \rightarrow & 0 \\ P & \rightarrow & 1 \\ P & \rightarrow & 0P0 \\ P & \rightarrow & 1P1 \end{array}$$

Context-Free Grammars

Definition (Context-Free Grammars)

A context-free grammar G is defined as a quadruple:

$$G = (V, T, S, P)$$

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

$$x \to y$$

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

Example: Palindromes

$$G = (\{P\}, \{0, 1\}, P, A)$$

where A is the set of five productions:

$$\begin{array}{lcl} P & \rightarrow & \epsilon \\ P & \rightarrow & 0 \\ P & \rightarrow & 1 \\ P & \rightarrow & 0P0 \\ P & \rightarrow & 1P1 \end{array}$$

Example: Simple Arithmetic Expressions

$$G = (\{E, I\}, \{+, *, (,), a, b, 0, 1\}, E, P)$$

where P is a set of productions:

$$\begin{array}{rcl} E & \rightarrow & I \\ E & \rightarrow & E + E \\ E & \rightarrow & E * E \\ E & \rightarrow & (E) \\ I & \rightarrow & a \\ I & \rightarrow & b \\ I & \rightarrow & Ia \\ I & \rightarrow & Ib \\ I & \rightarrow & I0 \\ I & \rightarrow & I0 \end{array}$$

Derivation

Definition (Derivation Relation, $\Rightarrow$)

Let $G = (V, T, S, P)$ be a context-free grammar. Let $\alpha A \beta$ be a string of terminals and variables, where $A \in V$ and $\alpha, \beta \in (V \cup T)^*$. Let $A \rightarrow \gamma$ is a production in G. Then, we say $\alpha A \beta$ derives $\alpha \gamma \beta$, and write

$$\alpha A \beta \Rightarrow \alpha \gamma \beta.$$

Definition ($\Rightarrow^*$, Closure of $\Rightarrow$)

$\Rightarrow^*$ is a relation that represents zero, or more steps of derivations:

- Basis: For any string α of terminals and variables, $\alpha \Rightarrow^* \alpha$.
- Induction: If $\alpha \Rightarrow^* \beta$ and $\beta \Rightarrow \gamma$, then $\alpha \Rightarrow^* \gamma$.

Example

A derivation for $a * (a + b00)$:

$$E \Rightarrow E * E \Rightarrow I * E \Rightarrow a * E \Rightarrow a * (E) \Rightarrow$$
$$a * (E + E) \Rightarrow a * (I + E) \Rightarrow a * (a + E) \Rightarrow a * (a + I) \Rightarrow$$
$$a * (a + I0) \Rightarrow a * (a + I00) \Rightarrow a * (a + b00)$$

Thus, $E \Rightarrow^* a * (a + b00)$.

Leftmost and Rightmost Derivations

- Leftmost derivation: replace the leftmost variable at each derivation step
- Rightmost derivation: replace the rightmost variable at each derivation step

The right most derivation for $a * (a + b00)$:

$$E \Rightarrow E * E \Rightarrow E * (E) \Rightarrow E * (E + E) \Rightarrow E * (E + I) \Rightarrow E * (E + I0) \Rightarrow E * (E + I00) \Rightarrow E * (E + b00) \Rightarrow E * (I + b00) \Rightarrow E * (a + b00) \Rightarrow I * (a + b00) \Rightarrow a * (a + b00)$$

Language of a Grammar

Definition

Let $G = (V, T, S, P)$ be a context-free grammar. The language of G, denoted $L(G)$, is the set of terminal strings that have derivations from the start symbol. That is,

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

Definition (Context-free Language)

If a language L is the language of some context-free grammar G, i.e., $L = L(G)$, then we say L is a context-free language, shortly CFL.

Sentential Forms

Definition (Sentential Forms)

If $G = (V, T, S, P)$ is a context-free grammar, then any string $\alpha \in (V \cup T)^*$ such that $S \Rightarrow^* \alpha$ is a *sentential form*.

- If $S \Rightarrow^* \alpha$ is a leftmost derivation, α is a *left-sentential form*.
- If $S \Rightarrow^* \alpha$ is a rightmost derivation, α is a *right-sentential form*.

Example

- Leftmost:

$$E \Rightarrow E * E \Rightarrow I * E \Rightarrow a * E \Rightarrow a * (E)$$
$$a * (E + E) \Rightarrow a * (I + E) \Rightarrow a * (a + E) \Rightarrow a * (a + I) \Rightarrow$$
$$a * (a + I0) \Rightarrow a * (a + I00) \Rightarrow a * (a + b00)$$

- Neither leftmost nor rightmost:

$$E \Rightarrow E * E \Rightarrow E * (E) \Rightarrow E * (E + E) \Rightarrow E * (I + E)$$

Exercises

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

Exercises

- $L = \{ww^R \mid w \in \{a, b\}^*\}$
- $L = \{a^n b^n \mid n \geq 0\}$

Exercises

- $L = \{ww^R \mid w \in \{a,b\}^*\}$
- $L = \{a^n b^n \mid n \geq 0\}$
- $L = \{a^n b^m \mid n \neq m\}$

Exercises

- $L = \{ww^R \mid w \in \{a, b\}^*\}$
- $L = \{a^n b^n \mid n \geq 0\}$
- $L = \{a^n b^m \mid n \neq m\}$
- The language of balanced parentheses.
  - E.g., ϵ, (), ()(), (()), (()())