Context-Free Grammars

Jiří Techet   Tomáš Masopust   Alexander Meduna

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Modern Formal Language Theory, 2007

Context-Free Grammar

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

- N alphabet of nonterminals
- T alphabet of terminals
- P finite set of productions of the form

$$A \rightarrow x$$

  with $A \in N$ and $x \in (N \cup T)^*$
- S the start symbol, $S \in N$

Proper Context-Free Grammar

Useful Symbol

A symbol $X \in N \cup T$ is useful if

1. $S \Rightarrow^* uXv$
2. $X \Rightarrow^* y$

for some $u, v \in (N \cup T)^*$ and $y \in T^*$

Proper Context-Free Grammar

A context-free grammar $G = (N, T, P, S)$ is proper if

1. $N \cup T$ contains only useful symbols
2. G is ε-free
3. G is unit-free

Properties of Proper Context-Free Grammars

Theorem

For every context-free language L, there is a proper context-free grammar G such that

$$L - \{\varepsilon\} = L(G)$$

Claim

If $G = (N, T, P, S)$ is proper, then for every $A \in N$

$$S \Rightarrow^* uAy \Rightarrow^* uwy$$

with $u, w, y \in T^*$

Weak Pumping Lemma

Weak Pumping Lemma

Let L be an infinite context-free language. Then, L contains a string $z = uvwxy$ such that

1. $uv^iwx^iy \in L$ for every $i \geq 0$
2. $|vx| \geq 1$

Weak Pumping Lemma – Proof

Let G be a proper context-free grammar such that $L = L(G)$

1. By contradiction: assume that no derivation in G contains two identical nonterminals. Then, $L(G)$ is finite – a contradiction.
2. Thus, there is

$$S \Rightarrow^* u'Ay' \Rightarrow^+ u'v'Ax'y' \Rightarrow^* u'v'wx'y'$$

in G, where $u', v', x', y' \in (N \cup T)^*$, $A \in N$, $w \in T^*$, $|v'x'| \geq 1$. As G is proper,

$$u' \Rightarrow^* u, v' \Rightarrow^* v, x' \Rightarrow^* x, \text{ and } y' \Rightarrow^* y$$

for some $u, v, x, y \in T^*$, $|vx| \geq 1$. Therefore,

$$S \Rightarrow^* uAy \Rightarrow^+ uvAxy \Rightarrow^* uvwxy.$$

Thus, $uv^iwx^iy \in L$ for every $i \geq 0$. □

Weak Pumping Lemma – Example

Example

Consider $L = \{a^n b^n c^n \ : \ n \geq 0\}$. By weak pumping lemma, L contains $z = uvwxy$ such that $|vx| \geq 1$ and $uv^i wx^i y \in L$ for every $i \geq 0$.

1. Let v or x be in
$$\{a\}^+\{b\}^+ \cup \{b\}^+\{c\}^+ \cup \{a\}^+\{b\}^+\{c\}^+.$$
Then, $uvvwxxy \notin L$ – contradiction.
2. Let v or x be in
$$\{a\}^+ \cup \{b\}^+ \cup \{c\}^+.$$
Then, $uwy \notin L$ – contradiction. □

Pumping Lemma

Pumping Lemma

Let L be a context-free language. Then, there is $k \geq 1$ such that for every $z \in L$ with $|z| \geq k$,

$$z = uvwxy$$

so that

1. $vx \neq \varepsilon$
2. $|vwx| \leq k$
3. $uv^m wx^m y \in L$ for all $m \geq 0$.

Pumping Lemma – Example

Example

Consider $L = \{a^{n^2} : n \geq 1\}$. Set $z = a^{k^2}$, where k is the pumping lemma constant. As $k^2 \geq k$, $|z| \geq k$. Express z as

$$z = uvwxy.$$

By pumping lemma, $uv^2wx^2y \in L$. Observe that $|vx| \leq k$, so

$$k^2 = |uvwxy| < |uv^2wx^2y| = |uvwxy| + |vx| \leq$$
$$k^2 + k < k^2 + 2k + 1 = (k+1)^2.$$

As $k^2 < |uv^2wx^2y| < (k+1)^2$, $uv^2wx^2y \notin L$ – contradiction. L is not a context-free language. □

Homework Assignment

1. Establish a pumping lemma for regular languages (based on regular grammars). Use this lemma to prove that some context-free languages are not regular.
2. By using this lemma, demonstrate that a computer program that decides whether a positive integer n is prime cannot be based on any finite automaton.

Bibliography

A. Meduna.
Automata and Languages: Theory and Applications.
Springer, London, 2000.

G. Rozenberg and A. Salomaa.
Handbook of Formal Languages, volume 1–3.
Springer, Berlin, 1997.

A. Salomaa.
Formal Languages.
Academic Press, New York, 1973.