Teil IV

Kontextfreie Sprachen, Teil 3: CYK-Algorithmus und Mehrdeutigkeit

Das Wortproblem für kontextfreie Grammatiken

Wortproblem für CFGs

Eingabe: CFG G, Wort w.
Frage: Ist $w \in \mathcal{L}(G)$?

Das Wortproblem für kontextfreie Grammatiken

Wortproblem für CFGs

Eingabe: CFG G, Wort w.
Frage: Ist $w \in \mathcal{L}(G)$?

Idee Nr. 1

Betrachte CFGs in CNF.

Das Wortproblem für kontextfreie Grammatiken

Wortproblem für CFGs

Eingabe: CFG G, Wort w.
Frage: Ist $w \in \mathcal{L}(G)$?

$$G := (\Sigma, V, P, S)$$

Chomsky-Normalform

$A \to BC$, $A \to \mathtt{b}$, $S \to \varepsilon$.

Idee Nr. 1

Betrachte CFGs in CNF.

Das Wortproblem für kontextfreie Grammatiken

Wortproblem für CFGs

Eingabe: CFG G, Wort w.
Frage: Ist $w \in \mathcal{L}(G)$?

$$G := (\Sigma, V, P, S)$$

Chomsky-Normalform

$A \to BC$, $A \to \mathtt{b}$, $S \to \varepsilon$.

Idee Nr. 1

Betrachte CFGs in CNF.

Idee Nr. 2

Zerlege w in Buchstaben:

$$w = a_1 \cdots a_n$$

Das Wortproblem für kontextfreie Grammatiken

Wortproblem für CFGs

Eingabe: CFG G, Wort w.
Frage: Ist $w \in \mathcal{L}(G)$?

$$G := (\Sigma, V, P, S)$$

Chomsky-Normalform

$A \to BC$, $A \to \mathtt{b}$, $S \to \varepsilon$.

Idee Nr. 1

Betrachte CFGs in CNF.

Idee Nr. 2

Zerlege w in Buchstaben:
$$w = a_1 \cdots a_n$$

Idee Nr. 3:

Berechne folgende Mengen:

$$V[i,j] := \{A \in V \mid A \Rightarrow_G^* a_i \cdots a_j\}$$

(mit $1 \leq i \leq j \leq n$, $n = |w|$)

Die Mengen $V[i, j]$

$$V[i, i] = \{A \mid A \to a_i \in P\},$$

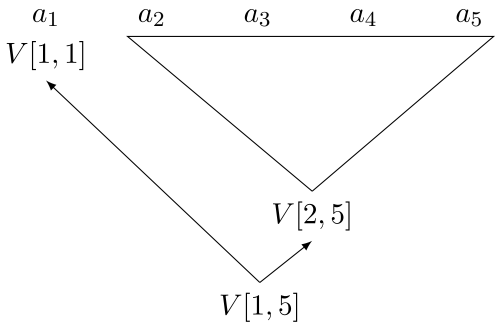
Die Mengen $V[i,j]$

$$V[i,i] = \{A \mid A \to a_i \in P\},$$

$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$

Die Mengen $V[i, j]$

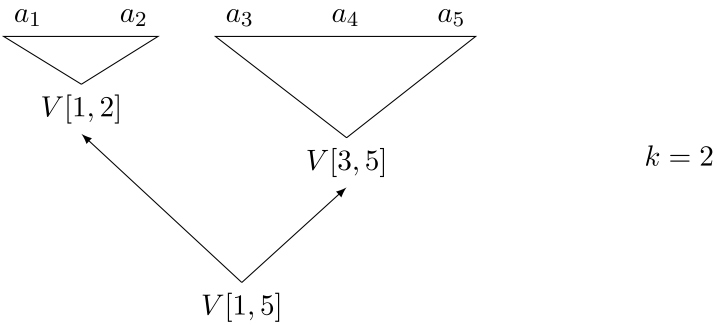
$$V[i,i] = \{A \mid A \to a_i \in P\},$$
$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$



$k = 1$

Die Mengen $V[i, j]$

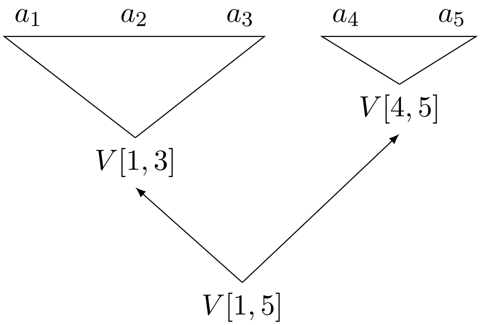
$$V[i,i] = \{A \mid A \to a_i \in P\},$$
$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$



Die Mengen $V[i,j]$

$$V[i,i] = \{A \mid A \to a_i \in P\},$$

$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$

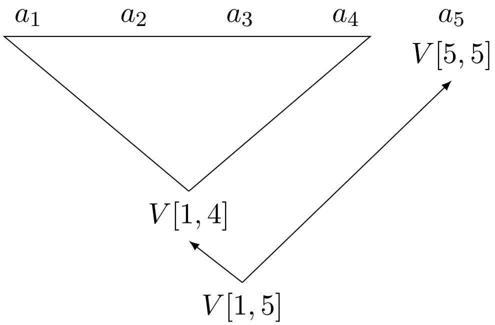


$k = 3$

Die Mengen $V[i,j]$

$$V[i,i] = \{A \mid A \to a_i \in P\},$$

$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$



$k = 4$

Der Cocke-Younger-Kasami-Algorithmus ($\mathtt{CYK}$)

Eingabe : Eine CFG $G := (\Sigma, V, P, S)$ in CNF,
Wort $w = a_1 \cdots a_n$, $n \geq 1$

Ausgabe : true, falls $w \in \mathcal{L}(G)$

for $i := 1$ **to** n **do** $V[i,i] := \{A \in V \mid A \to a_i$ ist Regel in $P\}$;

for $d := 1$ **to** $n-1$ **do**

  for $i := 1$ **to** $n-d$ **do**

    $j := i + d$;

    $V[i,j] := \emptyset$;

    for $k := i$ **to** $j-1$ **do**

      $V_{neu} := \{A \in V \mid A \to BC$ ist Regel in $P, B \in V[i,k],$
      $C \in V[k+1,j]\}$;

      $V[i,j] := V[i,j] \cup V_{neu}$;

if $S \in V[1,n]$ **then return true**;

else return false;

Analyse von CYK

Satz

Sei $G := (\Sigma, V, P, S)$ CFG in CNF, sei $w \in \Sigma^+$.

CYK entscheidet $w \in \mathcal{L}(G)$ in Zeit $O(mn^3)$, wobei $m := |P|$ und $n := |w|$.

Beispiel für CYK

$$\Sigma := \{\mathtt{a}, \mathtt{b}\},$$
$$G := (\Sigma, V, P, S),$$
$$V := \{S, A, B, C\}$$

Beispiel für CYK

$$\Sigma := \{\mathtt{a}, \mathtt{b}\},$$
$$G := (\Sigma, V, P, S),$$
$$V := \{S, A, B, C\}$$

$$:= \mathtt{babb}$$

$$P := \{S \to AB \mid BC,$$
$$A \to BC \mid \mathtt{a},$$
$$B \to CA \mid \mathtt{b},$$
$$C \to AB \mid \mathtt{a}\}.$$

Beispiel für CYK

$$\Sigma := \{\mathtt{a}, \mathtt{b}\},$$
$$G := (\Sigma, V, P, S),$$
$$V := \{S, A, B, C\}$$

$$w := \mathtt{babb}$$

$$\begin{aligned} P := \{S &\to AB \mid BC, \\ A &\to BC \mid \mathtt{a}, \\ B &\to CA \mid \mathtt{b}, \\ C &\to AB \mid \mathtt{a}\}. \end{aligned}$$

Beispiel für CYK

$$\Sigma := \{\mathtt{a}, \mathtt{b}\},$$
$$G := (\Sigma, V, P, S),$$
$$V := \{S, A, B, C\}$$

$$w := \mathtt{babb}$$

$$P := \{S \to AB \mid BC,$$
$$A \to BC \mid \mathtt{a},$$
$$B \to CA \mid \mathtt{b},$$
$$C \to AB \mid \mathtt{a}\}.$$

b	a	b	b
$V[1,1]$	$V[2,2]$	$V[3,3]$	$V[4,4]$
$V[1,2]$	$V[2,3]$	$V[3,4]$	
$V[1,3]$	$V[2,4]$		
$V[1,4]$			

Beispiel für CYK

$$\Sigma := \{\mathtt{a}, \mathtt{b}\},$$
$$G := (\Sigma, V, P, S),$$
$$V := \{S, A, B, C\}$$

$$w := \mathtt{babb}$$

$$
\begin{aligned}
P := \{S &\to AB \mid BC,\\
A &\to BC \mid \mathtt{a},\\
B &\to CA \mid \mathtt{b},\\
C &\to AB \mid \mathtt{a}\}.
\end{aligned}
$$

b	a	b	b
$V[1,1]$	$V[2,2]$	$V[3,3]$	$V[4,4]$
$V[1,2]$	$V[2,3]$	$V[3,4]$	
$V[1,3]$	$V[2,4]$		
$V[1,4]$			

$$V[i,i] = \{A \mid A \to a_i \in P\},$$
$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$

Syntaxanalyse

- CYK überprüft $w \in \mathcal{L}(G)$

Syntaxanalyse

- CYK überprüft $w \in \mathcal{L}(G)$
- Syntaxanalyse: berechne Ableitungsbaum für w in G

Syntaxanalyse

- CYK überprüft $w \in \mathcal{L}(G)$
- Syntaxanalyse: berechne Ableitungsbaum für w in G
- möglich durch kleine Modifikation von CYK:
  speichere in Tabelle anstelle von A 4-Tupel $A : B, k, C$

$$V[i,j] = \bigcup_{i \le k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$

Syntaxanalyse

- CYK überprüft $w \in \mathcal{L}(G)$
- Syntaxanalyse: berechne Ableitungsbaum für w in G
- möglich durch kleine Modifikation von CYK:
  speichere in Tabelle anstelle von A 4-Tupel $A : B, k, C$

$$V[i,j] = \bigcup_{i \leq k < j} \{A \mid A \to BC \in P, B \in V[i,k], C \in V[k+1,j]\}$$

- Aus Tabelleneinträgen können alle Ableitungsbäume für w rekonstruiert werden.

Mehrdeutige Grammatiken

Definition

Sei $G := (\Sigma, V, P, S)$ CFG.

- $w \in \mathcal{L}(G)$ ist **eindeutig**, wenn genau ein Ableitungsbaum für w in G existiert.

Mehrdeutige Grammatiken

Definition

Sei $G := (\Sigma, V, P, S)$ CFG.

- $w \in \mathcal{L}(G)$ ist **eindeutig**, wenn genau ein Ableitungsbaum für w in G existiert.
- $w \in \mathcal{L}(G)$ ist **mehrdeutig**, wenn mehr als ein Ableitungsbaum für w in G existiert.

Mehrdeutige Grammatiken

Definition

Sei $G := (\Sigma, V, P, S)$ CFG.

- $w \in \mathcal{L}(G)$ ist **eindeutig**, wenn genau ein Ableitungsbaum für w in G existiert.
- $w \in \mathcal{L}(G)$ ist **mehrdeutig**, wenn mehr als ein Ableitungsbaum für w in G existiert.

- G eindeutig: **jedes** $w \in \mathcal{L}(G)$ eindeutig

Mehrdeutige Grammatiken

Definition

Sei $G := (\Sigma, V, P, S)$ CFG.

- $w \in \mathcal{L}(G)$ ist **eindeutig**, wenn genau ein Ableitungsbaum für w in G existiert.
- $w \in \mathcal{L}(G)$ ist **mehrdeutig**, wenn mehr als ein Ableitungsbaum für w in G existiert.

- G eindeutig: **jedes** $w \in \mathcal{L}(G)$ eindeutig
- G mehrdeutig: mindestens ein $w \in \mathcal{L}(G)$ mehrdeutig

Linksableitungen

- $\gamma_i \Rightarrow_G \gamma_{i+1}$ ist **Linksableitungsschritt**:
  am weitesten links stehende Variable in γ_i wird ersetzt

Linksableitungen

- $\gamma_i \Rightarrow_G \gamma_{i+1}$ ist **Linksableitungsschritt**:
  am weitesten links stehende Variable in γ_i wird ersetzt
- **Linksableitung**: Ableitung aus Linksableitungsschritten

Linksableitungen

- $\gamma_i \Rightarrow_G \gamma_{i+1}$ ist **Linksableitungsschritt**:
  am weitesten links stehende Variable in γ_i wird ersetzt
- **Linksableitung**: Ableitung aus Linksableitungsschritten

- jede Linksableitung entspricht genau einem Ableitungsbaum

Linksableitungen

- $\gamma_i \Rightarrow_G \gamma_{i+1}$ ist **Linksableitungsschritt**:
  am weitesten links stehende Variable in γ_i wird ersetzt
- **Linksableitung**: Ableitung aus Linksableitungsschritten

- jede Linksableitung entspricht genau einem Ableitungsbaum
- jeder Ableitungsbaum entspricht genau einer Linksableitung

Mehrdeutigkeit und Sprachen

Definition

Eine kontextfreie Sprache L ist

- **inhärent mehrdeutig**, wenn jede CFG G für L mehrdeutig ist,

Mehrdeutigkeit und Sprachen

Definition

Eine kontextfreie Sprache L ist

- **inhärent mehrdeutig**, wenn jede CFG G für L mehrdeutig ist,
- **nicht inhärent mehrdeutig**, wenn mindestens eine eindeutige CFG G für L existiert.

Mehrdeutigkeit und Sprachen

Definition

Eine kontextfreie Sprache L ist

- **inhärent mehrdeutig**, wenn jede CFG G für L mehrdeutig ist,
- **nicht inhärent mehrdeutig**, wenn mindestens eine eindeutige CFG G für L existiert.

Lemma

Jede reguläre Sprache ist nicht inhärent mehrdeutig.

Mehrdeutigkeit und Sprachen

Definition

Eine kontextfreie Sprache L ist

- **inhärent mehrdeutig**, wenn jede CFG G für L mehrdeutig ist,
- **nicht inhärent mehrdeutig**, wenn mindestens eine eindeutige CFG G für L existiert.

Lemma

Jede reguläre Sprache ist nicht inhärent mehrdeutig.

Korollar

Jede unäre kontextfreie Sprache ist nicht inhärent mehrdeutig.

Inhärent mehrdeutige Sprachen

Satz

Es existieren kontextfreie Sprachen, die inhärent mehrdeutig sind.

Inhärent mehrdeutige Sprachen

Satz

Es existieren kontextfreie Sprachen, die inhärent mehrdeutig sind.

Beispiel 1

$$L_1 := \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i,j \in \mathbb{N}, i = j \text{ oder } j = k\}$$

Satz

Es existieren kontextfreie Sprachen, die inhärent mehrdeutig sind.

Beispiel 1

$$L_1 := \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i, j \in \mathbb{N}, i = j \text{ oder } j = k\}$$

Beispiel 2

$$\begin{aligned} L_2 := &\{\mathtt{a}^m\mathtt{b}^m\mathtt{a}^n\mathtt{b}^n \mid m, n \in \mathbb{N}_{>0}\} \\ &\cup \{\mathtt{a}^m\mathtt{b}^n\mathtt{a}^n\mathtt{b}^m \mid m, n \in \mathbb{N}_{>0}\}. \end{aligned}$$