

Theoretische Grundlagen der Informatik

Vorlesung am 12.01.2012



Syntaxbäume



Syntaxbäume visualisieren die Ableitung eines einzelnen Wortes.

- An der Wurzel eines Syntaxbaumes steht das Startsymbol.
- Jeder innere Knoten enthält eine Variable.
- Die Blätter sind Symbole aus Σ oder ε .
- Wenn ein innerer Knoten A als Nachfolger von links nach rechts $\alpha_1, \ldots, \alpha_r \in V \cup \Sigma$ hat, so muss $A \to \alpha_1 \ldots \alpha_r$ eine Ableitungsregel der Grammatik sein.

Syntaxbäume - Beispiel

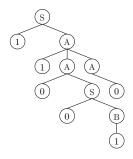


Zu den Regeln

$$R = \{S \rightarrow 0B \mid 1A, A \rightarrow 0 \mid 0S \mid 1AA, B \rightarrow 1 \mid 1S \mid 0BB\}$$

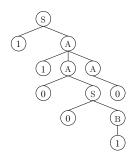
betrachte die Ableitung

$$S
ightarrow 1A
ightarrow 11AA
ightarrow 11A0
ightarrow 110S0
ightarrow 1100B0
ightarrow 110010$$



Syntaxbäume - Beispiel





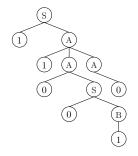
- Zu jeder Ableitung gehört genau ein Syntaxbaum
- Zu jedem Syntaxbaum gehören jedoch verschiedene Ableitungen des gleichen Wortes.

Syntaxbäume - Beispiel



$$R = \{S \to 0B \mid 1A, A \to 0 \mid 0S \mid 1AA, B \to 1 \mid 1S \mid 0BB\}$$

 $S \to 1A \to 11AA \to 110SA \to 1100BA \to 11001A \to 110010$
 $S \to 1A \to 11AA \to 11A0 \to 110S0 \to 1100B0 \to 110010$



Links/Rechtsableitung, Eindeutigkeit



Für Chomsky-2 Grammatiken ailt:

 Wegen der Kontextfreiheit ist die Reihenfolge, in der abgeleitet wird, für das Ergebnis unerheblich.

Eine **Linksableitung** (**Rechtsableitung**) ist eine Ableitung, bei der in jedem Schritt die linkeste (rechteste) Variable abgeleitet wird.

Eine kontextfreie Grammatik G heißt eindeutig, wenn es für jedes Wort $w \in L(G)$ genau einen Syntaxbaum gibt.

Eine kontextfreie Sprache L heißt eindeutig, wenn es eine eindeutige Grammatik G mit L(G) = L gibt. Ansonsten heißt L inhärent mehrdeutig.

Beispiel



Die Sprache

$$L = \{0^n 1^n \mid n \ge 1\}$$

erzeugt durch die Grammatik

$$V = \{S\}$$

 $\Sigma = \{0, 1\}$
 $R = \{S \rightarrow 01 \mid 0S1\}$.

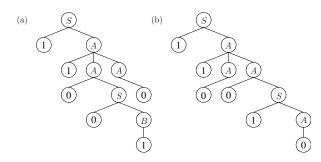
ist eindeutig.



Die Sprache gegeben durch die Regeln

$$R = \{S \rightarrow 0B \mid 1A, A \rightarrow 0 \mid 0S \mid 1AA, B \rightarrow 1 \mid 1S \mid 0BB\}$$

ist nicht eindeutig.



Chomsky-Normalform



Eine kontextfreie Grammatik ist in **Chomsky–Normalform**, wenn alle Regeln von der Form:

$$A \rightarrow BC$$
 oder $A \rightarrow a$

sind, mit A, B, $C \in V$ und $a \in \Sigma$.



$$A \rightarrow BC$$
 oder $A \rightarrow a$

- \blacksquare Grammatiken in Chomsky–Normalform können also nicht das Wort ε erzeugen.
- ullet Für kontextfreie Sprachen, die arepsilon enthalten, läßt sich eine Grammatik leicht ergänzen durch die Regeln

$$\mathcal{S}'
ightarrow arepsilon \qquad \qquad \mathcal{S}'
ightarrow \mathcal{S}$$

wobei S' ein neues Startsymbol zur Erzeugung von ε ist.



Satz:

Jede kontextfreie Grammatik, die nicht das leere Wort erzeugt, kann in eine Grammatik in Chomsky-Normalform überführt werden.

Beweis (konstruktiv):

- Wir geben eine Schritt-für-Schritt-Überführung der Regeln in Regeln in Normalform an.
- Großbuchstaben repräsentieren immer Nichtterminale
- Kleinbuchstaben repräsentieren immer Terminale



Wir veranschaulichen den Beweis an folgendem Beispiel: Grammatik $G = (\Sigma, V, S, R)$ mit

$$\Sigma = \{a, b\}$$

$$V = \{A, B, C, D, E, S\}$$

und der folgenden Regelmenge R:

$$S \rightarrow A \mid aAa \mid bBb \mid \varepsilon$$
 $A \rightarrow C \mid a$
 $B \rightarrow b$
 $C \rightarrow AF \mid CDE \mid \varepsilon$
 $D \rightarrow A \mid B \mid ab$
 $E \rightarrow B$
 $F \rightarrow D \mid E$

Schritt 1:



Ziel:

• Alle Regeln enthalten auf der rechten Seite nur Symbole aus V oder nur ein Symbol aus Σ .

Vorgehen:

Ersetze dazu in allen rechten Seiten von Regeln Symbole aus $a \in \Sigma$ durch neue Variablen Y_a und füge die Regeln $Y_a \to a$ hinzu.

Schritt 1



$$S \rightarrow A \mid aAa \mid bBb \mid \varepsilon$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid CDE \mid \varepsilon$$

$$D \rightarrow A \mid B \mid ab$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$S \rightarrow A \mid Y_{a}AY_{a} \mid Y_{b}BY_{b} \mid \varepsilon$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid CDE \mid \varepsilon$$

$$D \rightarrow A \mid B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$Y_{a} \rightarrow a$$

$$Y_{b} \rightarrow b$$

Schritt 2



Ziel:

■ Alle rechten Seiten haben Länge ≤ 2.

Vorgehen:

- Sei $A \rightarrow B_1 \dots B_m$ Regel mit m > 2.
- Führe m-2 neue Variablen C_1, \ldots, C_{m-2} ein, und ersetze die Regel

$$A \rightarrow B_1 \dots B_m$$

durch neue Regeln

$$egin{array}{cccc} A &
ightarrow & B_1C_1 \ C_i &
ightarrow & B_{i+1}C_{i+1} \ C_{m-2} &
ightarrow & B_{m-1}B_m \end{array}$$
 für $1 \leq i \leq m-3$

Schritt 2



$$S \rightarrow A \mid Y_{a}AY_{a} \mid Y_{b}BY_{b} \mid \varepsilon$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid CDE \mid \varepsilon$$

$$D \rightarrow A \mid B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$Y_a \rightarrow a$$

 $Y_b \rightarrow b$

$$S \rightarrow A \mid Y_{a}C_{1} \mid Y_{b}C_{2} \mid \varepsilon$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid CC_{3} \mid \varepsilon$$

$$D \rightarrow A \mid B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$C_{1} \rightarrow AY_{a}$$

$$C_{2} \rightarrow BY_{b}$$

$$C_{3} \rightarrow DE$$

$$Y_{a} \rightarrow a$$

$$Y_{b} \rightarrow b$$

Schritt 3:



Ziel:

Es kommen keine Regeln $A \rightarrow \varepsilon$ vor.

Vorgehen, Phase 1:

Finde die Menge V' aller Variablen A für die $A \stackrel{*}{\to} \varepsilon$ existiert:

- **E**s werden erst alle A mit $A \rightarrow \varepsilon$ in V' aufgenommen.
- Dann wird geprüft, ob neue Regeln $B \to \varepsilon$ entstehen, wenn man A in allen Regeln auf der rechten Seite A durch ε ersetzt.
- Ist dies der Fall, so werden die entsprechenden Variablen B in V' aufgenommen und genauso behandelt.
- lacktriangle Das Verfahren hört auf, wenn V' sich nicht mehr ändert

Schritt 3:



Ziel:

Es kommen keine Regeln $A \rightarrow \varepsilon$ vor.

Vorgehen, Phase 1:

Finde die Menge V' aller Variablen A für die $A\stackrel{*}{ o} \varepsilon$ existiert:

- Bemerkung: In Phase 1 werden noch keine Regeln geändert
- Die Ersetzung ist also nur "testweise"
- Am Ende enthält V' alle Variablen A mit $A\stackrel{*}{ o} arepsilon.$

Schritt 3:



Ziel:

Es kommen keine Regeln $A \rightarrow \varepsilon$ vor.

Vorgehen, Phase 2: Ersetzung.

- Gegeben V' aus Phase 1
- Streiche alle Regeln $A \to \varepsilon$
- Für A → BC füge die zusätzliche Regel
 - $A \rightarrow B$ falls $C \in V'$
 - lacksquare A o C falls $B \in V'$

ein.

• (Die Regel $A \rightarrow BC$ wird nicht gestrichen).



Initialisierung

$$V' = \{A \mid A \rightarrow \varepsilon\} = \{S, C\}$$

- Erster Durchlauf liefert: A⇒ $V' = \{A, C, S\}$
- Zweiter Durchlauf liefert: D⇒ $V' = \{A, C, D, S\}$
- Dritter Durchlauf liefert: $F \Rightarrow V' = \{A, C, D, F, S\}$
- Vierter Durchlauf liefert nichts neues

$$S \rightarrow A \mid Y_a C_1 \mid Y_b C_2 \mid \varepsilon$$

$$A \rightarrow C \mid a$$

$$\mathsf{B} \; o \; \mathsf{b}$$

$$C \rightarrow AF \mid CC_3 \mid \varepsilon$$

$$D \rightarrow A \mid B \mid Y_a Y_b$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$C_1 \rightarrow AY_a$$

$$C_2 \rightarrow BY_b$$

$$C_3 \rightarrow DE$$

$$Y_a \rightarrow a$$

$$Y_b \rightarrow k$$



- Streiche ε-Produktionen
- Simuliere Ableitungen A ^{*}→ ε auf den verbleibenden Regeln
 V' = {A, C, D, F, S}

```
S \rightarrow A \mid Y_a C_1 \mid Y_b C_2
 A \rightarrow C \mid a
 B \rightarrow b
 C \rightarrow AF \mid F \mid A \mid CC_3 \mid C_3
 D \rightarrow A \mid B \mid Y_a Y_b
 F \rightarrow D \mid E
C_1 \rightarrow AY_a \mid Y_a
C_2 \rightarrow BY_h
C_3 \rightarrow DE \mid E
Y_a \rightarrow a
```

Schritt 4.



Ziel

- lacktriangle Die Grammatik enthält keine (Ketten-)Regeln der Form $A \to B$.
- Beispiel

$$A \rightarrow B \mid C$$

$$B \rightarrow C$$

$$C \rightarrow c$$

Schritt 4.



Ziel

Die Grammatik enthält keine (Ketten-)Regeln der Form $A \rightarrow B$.

Vorgehen

- **Phase 1:** Finde alle Kreise $A_1 \rightarrow A_2 \rightarrow ... \rightarrow A_n \rightarrow A_1$ Ersetze alle A_i durch A_1
- Phase 2: Betrachte die Regeln der Form A → B in umgekehrt topologischer Reihenfolge
 - Für Regel $A \rightarrow B$ und jede Regel $B \rightarrow b$ füge Regel $A \rightarrow b$ hinzu
 - lacktriangle Lösche Regel A o B

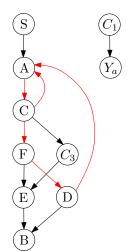


Topologische Sortierung der Regelmenge

- V_1, \ldots, V_k Menge von Variablen aus Kettenregeln
- V_1, \ldots, V_k topologisch sortiert, wenn gilt:
- $V_i \stackrel{*}{\to} V_j \Rightarrow i < j$
- Voraussetzung: Es gibt keine zyklischen Abhängigkeiten

Abhängigkeitsgraph





$$S \rightarrow A \mid Y_{a}C_{1} \mid Y_{b}C_{2}$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid F \mid A \mid CC_{3} \mid C_{3}$$

$$D \rightarrow A \mid B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$C_{1} \rightarrow AY_{a} \mid Y_{a}$$

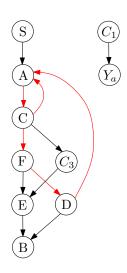
$$C_{2} \rightarrow BY_{b}$$

$$C_{3} \rightarrow DE \mid E$$

$$Y_{a} \rightarrow a$$

Schritt 4 - Phase 1





- Zyklen $A \rightarrow C$ und $A \rightarrow C \rightarrow F \rightarrow D \rightarrow A$
- \Rightarrow A, C, F, D äquivalent
- Entferne an Zyklen beteiligte Regeln
- Ersetze Vorkommen von C, F, D in allen Regeln durch A
- Lösche Regeln der Form $A \rightarrow A$

Zyklus:
$$A \rightarrow C \rightarrow F \rightarrow D \rightarrow A$$



$$S \rightarrow A \mid Y_{a}C_{1} \mid Y_{b}C_{2}$$

$$A \rightarrow C \mid a$$

$$B \rightarrow b$$

$$C \rightarrow AF \mid F \mid A \mid CC_{3} \mid C_{3}$$

$$D \rightarrow A \mid B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$F \rightarrow D \mid E$$

$$C_{1} \rightarrow AY_{a} \mid Y_{a}$$

$$C_{2} \rightarrow BY_{b}$$

$$C_{3} \rightarrow DE \mid E$$

$$Y_{a} \rightarrow a$$

$$S \rightarrow A \mid Y_{a}C_{1} \mid Y_{b}C_{2}$$

$$A \rightarrow a$$

$$B \rightarrow b$$

$$A \rightarrow AA \mid AC_{3} \mid C_{3}$$

$$A \rightarrow B \mid Y_{a}Y_{b}$$

$$E \rightarrow B$$

$$A \rightarrow E$$

$$C_{1} \rightarrow AY_{a} \mid Y_{a}$$

$$C_{2} \rightarrow BY_{b}$$

$$C_{3} \rightarrow AE \mid E$$

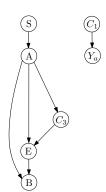
$$Y_{a} \rightarrow a$$

$$Y_{b} \rightarrow b$$

 $Y_b \rightarrow b$

Schritt 4 - Phase 2





$$S \rightarrow A \mid Y_a C_1 \mid Y_b C_2$$

$$A \rightarrow a \mid AA \mid AC_3 \mid C_3 \mid B \mid Y_a Y_b \mid E$$

$$B \rightarrow b$$

$$E \rightarrow B$$

$$C_1 \rightarrow AY_a \mid Y_a$$

$$C_2 \rightarrow BY_b$$

$$C_3 \rightarrow AE \mid E$$

$$Y_a \rightarrow a$$

$$Y_b \rightarrow b$$

Topologische Sortierung: S, A, C₃, E, B, C₁, Y_a



- Gehe in umgekehrter topologischer Sortierung vor
- Ersetze A o B durch A o eta, falls Regel B o eta existiert
- Topologische Sortierung: S, A, C₃, E, B, C₁, Y_a

$$S \rightarrow Y_a C_1 \mid Y_b C_2 \mid a \mid AA$$

$$AC_3 \mid AE \mid b \mid Y_a Y_b$$

$$A \rightarrow a \mid AA \mid AC_3 \mid AE \mid$$

$$b \mid Y_a Y_b$$

$$B \rightarrow b$$

$$E \rightarrow b$$

$$C_1 \rightarrow AY_a \mid a$$

$$C_2 \rightarrow BY_b$$

$$C_3 \rightarrow AE \mid b$$

$$Y_a \rightarrow a$$

$$Y_b \rightarrow b$$



Sonderbehandlung von ε -Produktionen

- Grammatik ist nun in Chomsky-Normalform, aber
- G enthält Regel $S \to \varepsilon$ (haben wir entfernt)
- Erweitere Grammatik um Regeln $S' \to S$ und $S' \to \varepsilon$ für neues Startsymbol S'

Der CYK-Algorithmus



Satz:

Es gibt einen Algorithmus (den Cocke-Younger-Kasami Algorithmus), der für eine kontextfreie Grammatik G in Chomsky-Normalform und ein Wort $w \in \Sigma^*$ in Zeit $\mathcal{O}(|R| \cdot n^3)$ entscheidet, ob $w \in L(G)$, wobei n = |w| und |R| die Anzahl der Regeln von G ist.

Beweis - Beschreibung des CYK-Algorithmus



- \blacksquare Sei $w = w_1 \dots w_n$.
- Sei $V_{ij} \subseteq V$ so dass $A \stackrel{*}{\rightarrow} w_i \dots w_j$ impliziert $A \in V_{ij}$.
- Für alle $1 \le i \le j \le n$ berechne die Menge $V_{ij} \subseteq V$.
- Dann ist $w \in L(G)$ genau dann, wenn $S \in V_{1n}$ ist.
- Die Tabelle der V_{ij} wird nach wachsendem $\ell := j i$ aufgebaut, beginnend mit $\ell = 0$.
- Für $j-i=\ell>0$ wird die Berechnung von V_{ij} systematisch auf zuvor berechnete V_{ik} , V_{k+1j} mit $i\leq k< j$ zurückgeführt

Bemerkung

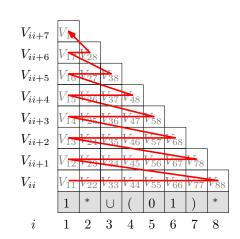
Wir benutzen dynamische Programmierung.



$$w = 1^* \cup (01)^*$$



$$w = 1^* \cup (01)^*$$



27



Fall $\ell = 0$:

- Nonstruiere die Mengen V_{ii} , d.h. alle $A \in V$ mit $A \stackrel{*}{ o} w_i$.
- Da G in Chomsky–Normalform ist, gilt $A \stackrel{*}{\rightarrow} w_i$ nur, wenn $(A \rightarrow w_i) \in R$.
- Die Berechnung von V_{ii} ist für alle $i \in \{1, ... n\}$ in $\mathcal{O}(|R|)$ möglich.



$$w = 1^* \cup (01)^*$$



$$w = 1^* \cup (01)^*$$

V_{ii+7}	V_{18}							
V_{ii+6}	V_{17}	V_{28}		,				
V_{ii+5}	V_{16}	V_{27}	V_{38}		,			
V_{ii+4}	V_{15}	V_{26}	V_{37}	V_{48}		,		
V_{ii+3}	V_{14}	V_{25}	V_{36}	V_{47}	V_{58}			
V_{ii+2}	V_{13}	V_{24}	V_{35}	V_{46}	V_{57}	V_{68}		
V_{ii+1}	V_{12}	V_{23}	V_{34}	V_{45}	V_{56}	V_{67}	V_{78}	
V_{ii}	S	Y_*	Y_{\cup}	$Y_{(}$	S	S	$Y_{)}$	Y_*
	1	*	Ů	(0	1)	*
i	1	2	3	4	5	6	7	8



Fall $\ell > 0$:

■ Jede Ableitung von $w_i \dots w_j$ muss mit einer Regel der Form

$$A \rightarrow BC$$

beginnen, wobei ein $k \in \{i, ..., j-1\}$ existiert mit

- lacksquare $B\stackrel{*}{ o} w_i \dots w_k$ und
- $C \stackrel{*}{\rightarrow} w_{k+1} \dots w_j.$



Verfahren

- Speichere alle Mengen V_{rs} als Arrays der Länge |V|, in denen für jedes $A \in V$ markiert ist, ob $A \in V_{rs}$.
- Berechnung von V_{ij} :
 Überprüfe für jede Regel $(A \rightarrow BC) \in R$ und jedes k, ob

$$B \stackrel{*}{\rightarrow} w_i \dots w_k$$

$$C \stackrel{*}{\rightarrow} w_{k+1} \dots w_j$$

durch Ansehen der Stelle

- B im Array zu V_{ik} und
- C im Array zu V_{k+1} j.



Verfahren

- Speichere alle Mengen V_{rs} als Arrays der Länge |V|, in denen für jedes $A \in V$ markiert ist, ob $A \in V_{rs}$.
- Berechnung von V_{ij} :
 Überprüfe für jede Regel $(A \rightarrow BC) \in R$ und jedes k, ob

$$B \stackrel{*}{\rightarrow} w_i \dots w_k$$

$$C \stackrel{*}{\rightarrow} w_{k+1} \dots w_j$$

durch Ansehen der Stelle

- B im Array zu V_{ik} und
- C im Array zu V_{k+1} j.

Bemerkung

Dies benötigt Aufwand in $\mathcal{O}(n \cdot |R|)$



$$w = 1^* \cup (01)^*$$

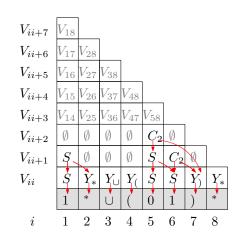


$$w = 1^* \cup (01)^*$$

V_{ii+7}	V_{18}							
V_{ii+6}	V_{17}	V_{28}		,				
V_{ii+5}	V_{16}	V_{27}	V_{38}					
V_{ii+4}	V_{15}	V_{26}	V_{37}	V_{48}				
V_{ii+3}	V_{14}	V_{25}	V_{36}	V_{47}	V_{58}			
V_{ii+2}	V_{13}	V_{24}	V_{35}	V_{46}	V_{57}	V_{68}		
V_{ii+1}	S_{\downarrow}	Ø	Ø	Ø	S_{\downarrow}	C_2	Ø	
V_{ii}	\dot{S}	Y_*	Y_{\cup}	$Y_{(}$	\dot{S}	\dot{S}	$Y_{)}$	Y_*
	1	*	Ů	(0	1)	*
i	1	2	3	4	5	6	7	8



$$w = 1^* \cup (01)^*$$



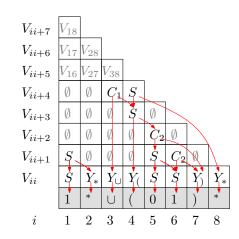


$$w = 1^* \cup (01)^*$$

V_{ii+7}	V_{18}							
V_{ii+6}	V_{17}	V_{28}						
V_{ii+5}	V_{16}	V_{27}	V_{38}					
V_{ii+4}	V_{15}	V_{26}	V_{37}	V_{48}				
V_{ii+3}	Ø	Ø	Ø	S_{\downarrow}	Ø			
V_{ii+2}	Ø	Ø	Ø	Ø	C_2	Ø		
V_{ii+1}	S_{\downarrow}	Ø	Ø	Ø	\dot{S}	C_2	Ø	
V_{ii}	\dot{S}	Y_*	Y_{\cup}	$Y_{(}$	\dot{S}	\dot{S}	$Y_{)}$	Y_*
	1	*	Ů		0	1)	*
i	1	2	3	4	5	6	7	8

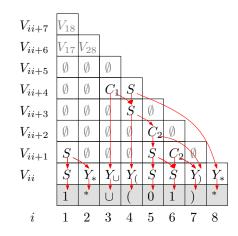


$$w = 1^* \cup (01)^*$$



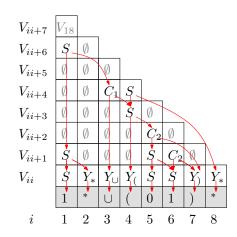


$$w = 1^* \cup (01)^*$$



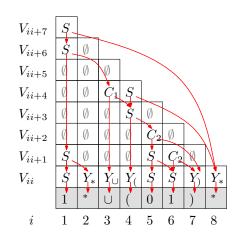


$$w = 1^* \cup (01)^*$$





$$w = 1^* \cup (01)^*$$



Ergebnisse zum Wortproblem



- Typ-0 Grammatik. Das Wortproblem ist nicht entscheidbar.
- Typ-1 Grammatik. Das Wortproblem ist NP-vollständig.
- Typ-2 Grammatik. Das Wortproblem ist in polynomieller Zeit lösbar.
- **Typ-3 Grammatik.** Das Wortproblem ist in linearer Zeit lösbar.