Inhaltsangabe

Dresden

§ 9. Abschlußeigenschaften kontextfreier Sprachen

Satz 9.1 (Abschlußeigenschaften)

Die Klasse $\mathcal{L}_2$ der kontextfreien Sprachen ist unter Vereinigung, Konkatenation und Kleene-Stern abgeschlossen.

Beweis:

Es sei $L_1 = L(G_1)$ und $L_2 = L(G_2)$ für kontextfreie Grammatiken $G_i = (N_i, \Sigma, P_i, S_i)$ $(i = 1, 2)$.

O.B.d.A. nehmen wir an, daß $N_1 \cap N_2 = \emptyset$.

1. $G := (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \longrightarrow S_1, S \longrightarrow S_2\}, S)$ mit $S \notin N_1 \cup N_2$ ist eine kontextfreie Grammatik mit $L(G) = L_1 \cup L_2$.

2. $G' := (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \longrightarrow S_1 S_2\}, S)$ mit $S \notin N_1 \cup N_2$ ist eine kontextfreie Grammatik mit $L(G') = L_1 \cdot L_2$.

3. $G'' := (N_1 \cup \{S\}, \Sigma, P_1 \cup \{S \longrightarrow \varepsilon, S \longrightarrow SS_1\}, S)$ mit $S \notin N_1$ ist eine kontextfreie Grammatik für L_1^*.

Wir werden zeigen: Abschluß unter Durchschnitt und Komplement nicht gilt.

Wie zeigt man das?

Angabe von Gegenbeispielen:

z.B. kontextfreie Sprachen L_1, L_2 angeben mit $L_1 \cap L_2$ nicht kontextfrei.

Wie zeigt man, daß ein Sprache nicht kontextfrei ist?

Pumping-Lemma

Der Beweis des Pumping-Lemmas für kontextfreie Sprachen verwendet den Begriff des Ableitungsbaums.

Beispiel 9.2

$$P = \{S \longrightarrow SbS,\ \ S \longrightarrow a\}$$

Drei Ableitungen von $ababa$:

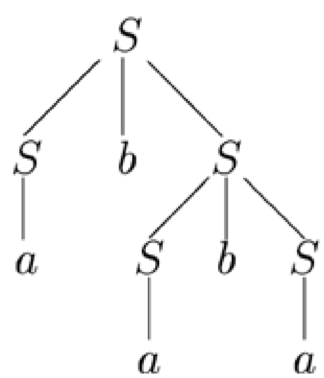
1. $S \vdash SbS \vdash abS \vdash abSbS \vdash ababS \vdash ababa$

2. $S \vdash SbS \vdash abS \vdash abSbS \vdash abSba \vdash ababa$

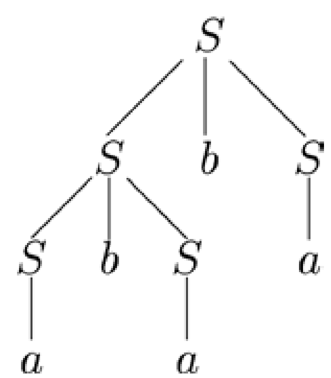
3. $S \vdash SbS \vdash Sba \vdash SbSba \vdash Sbaba \vdash ababa$

Die zugehörigen Ableitungsbäume:

Für (1) und (2):



Für (3):



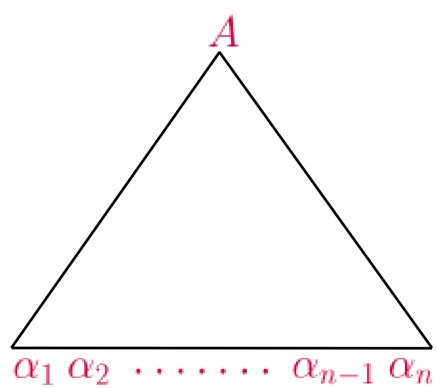
Ein Ableitungsbaum kann also für mehr als eine Ableitung stehen und dasselbe Wort kann verschiedene Ableitungsbäume haben.

Allgemein:

Die Knoten des Ableitungsbaumes sind mit Elementen aus $\Sigma \cup N$ beschriftet.

Ein mit A beschrifteter Knoten kann mit $\alpha_1, \ldots, \alpha_n$ beschriftete Nachfolgerknoten haben, wenn $A \longrightarrow \alpha_1 \ldots \alpha_n \in P$ ist.

Ein Ableitungsbaum, dessen Wurzel mit A beschriftet ist und dessen Blätter (von links nach rechts) mit $\alpha_1, \ldots, \alpha_n \in \Sigma \cup N$ beschriftet sind, beschreibt eine Ableitung $A \vdash_G^* \alpha_1 \ldots \alpha_n$.



$A \vdash_G^* \alpha_1 \ldots \alpha_n.$

Lemma 9.3 (Pumping-Lemma für kontextfreie Sprachen)

Es sei G eine kontextfreie Grammatik, die

- ε-frei ist,
- keine Regeln der Form $A \longrightarrow B$ enthält,
- m nichtterminale Symbole enthält,
- nur rechte Regelseiten der Länge $\leq k$ hat.

Es sei $n = k^{m+1}$.

Dann gibt es für jedes $z \in L(G)$ mit $|z| > n$ eine Zerlegung

$$z = uvwxy$$

mit:

- $vx \neq \varepsilon$ und $|vwx| \leq n$,
- $uv^iwx^iy \in L(G)$ für alle $i \geq 0$.

Dresden

Beweis:

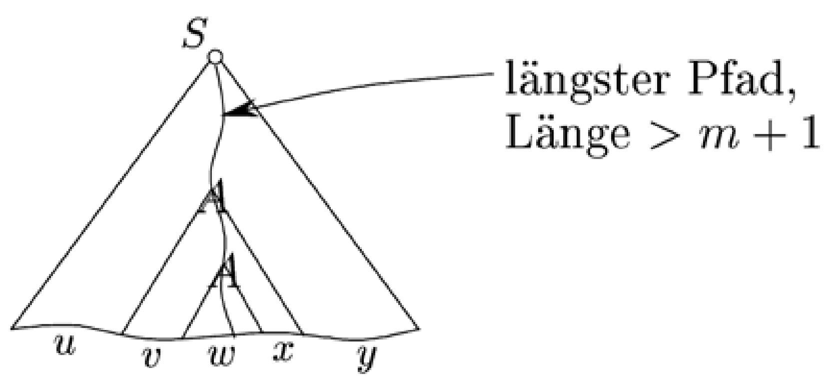
(1)

Ein Baum der Tiefe $\leq t$ und der Verzweigungszahl $\leq k$ hat maximal k^t viele Blätter.

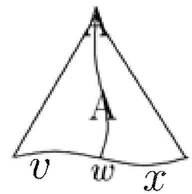
Da der Ableitungsbaum für z als Blattanzahl $|z| > k^{m+1}$ hat, ist die maximale Tiefe (längster Pfad von Wurzel bis Blatt) $> m + 1$.

(2)

Da es nur m verschiedene Elemente in N gibt, kommt auf diesem längsten Pfad ein nichtterminales Symbol A zweimal vor.



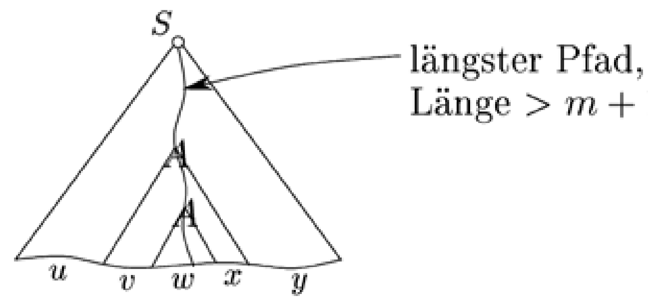
Wir wählen hier o.B.d.A. A so, daß es in dem Baum



keine andere Wiederholung von Nichterminalsymbolen gibt.

Daher hat dieser Baum eine Tiefe $\leq m+1$, was $|vwx| \leq k^{m+1} = n$ zeigt.

(3) Es gilt:



$$S \vdash_G^* uAy$$
$$A \vdash_G^* vAx$$
$$A \vdash_G^* w$$

woraus folgt: $S \vdash_G^* uAy \vdash_G^* uv^iAx^iy \vdash_G^* uv^iwx^iy$

(4) $vx \neq \varepsilon$:

Da G ε-frei ist, wäre sonst $A \vdash_G^+ vAx$ nur bei Anwesenheit von Regeln der Form $A \longrightarrow B$ möglich.

Dresden

Satz 9.4

$$\boxed{\mathcal{L}_2 \subset \mathcal{L}_1.}$$

Beweis:

Wir haben bereits gezeigt, daß $\mathcal{L}_2 \subseteq \mathcal{L}_1$ gilt.

Um zu beweisen, daß die Inklusion echt ist, zeigen wir

$$L := \{a^n b^n c^n \mid n \geq 1\} \in \mathcal{L}_1 \setminus \mathcal{L}_2.$$

(1) $L \notin \mathcal{L}_2$:

Angenommen, $L \in \mathcal{L}_2$.

Dann gibt es eine ε-freie kontextfreie Grammatik G ohne Regeln der Form $A \longrightarrow B$ für L.

Es sei $n_0 = k^{m+1}$ die zugehörige Zahl aus Lemma 9.3.

Wir betrachten $z := a^{n_0} b^{n_0} c^{n_0} \in L = L(G)$.

Dresden

Mit Lemma 9.3 gibt es eine Zerlegung von $z = a^{n_0}b^{n_0}c^{n_0}$:

$z = uvwxy,\ \ vx \neq \varepsilon$ und $uv^iwx^iy \in L$ für alle $i \geq 0$.

1. Fall:

v enthält verschiedene Buchstaben.

2. Fall:

x enthält verschiedene Buchstaben.

Führt zu entsprechendem Widerspruch.

3. Fall:

v enthält lauter gleiche Buchstaben und

x enthält lauter gleiche Buchstaben.

(2) $L \in \mathcal{L}_1$:

In Beispiel 6.3 hatten wir eine Grammatik G mit $L(G) = \{a^n b^n c^n \mid n \geq 1\}$ angegeben:

$$G = (N, \Sigma, P, S) \text{ mit } N = \{S, B\}, \Sigma = \{a, b, c\} \text{ und}$$
$$P = \{S \longrightarrow aSBc,\ S \longrightarrow abc,\ \underline{cB \longrightarrow Bc},\ bB \longrightarrow bb\}$$

nicht kontextsensitiv

Wir modifizieren G wie folgt:

- Ersetze in allen Produktionen c durch ein neues nichtterminales Symbol C und nimm die Produktion $C \longrightarrow c$ hinzu:

  $$\{S \longrightarrow aSBC,\ S \longrightarrow abC,\ CB \longrightarrow BC,\ bB \longrightarrow bb,\ C \longrightarrow c\}.$$

- Ersetze $CB \longrightarrow BC$ durch die kontextsensitiven Produktionen

  $$CB \longrightarrow A_1B,\ A_1B \longrightarrow A_1A_2,\ A_1A_2 \longrightarrow BA_2,\ BA_2 \longrightarrow BC.$$

  Diese Produktionen können nur dazu verwendet werden, $CB \longrightarrow BC$ zu simulieren.

Dresden

Beachte:

Auf diese Art kann man leicht zeigen, daß jede nichtkürzende Produktion

$$u \longrightarrow v \text{ mit } |u| \leq |v|$$

durch kontextsensitive Produktionen ersetzt werden kann, ohne die Sprache zu ändern.

Dies zeigt:

Die kontextsensitiven Sprachen sind genau die Sprachen, die durch nichtkürzende Grammatiken erzeugt werden, d.h. Grammtiken, deren Produktionen die folgende Form haben:

- $u \longrightarrow v$ mit $|u| \leq |v|$, oder
- $S \longrightarrow \varepsilon$ und S kommt nicht auf der rechten Seite einer Produktion vor.

Dresden

Korollar 9.5

Die Klasse $\mathcal{L}_2$ der kontextfreien Sprachen ist nicht unter Durchschnitt und Komplement abgeschlossen.

Beweis:

(1) Die Sprachen

$$\begin{aligned}\{a^n b^n c^m \mid n \geq 1, m \geq 1\} &= \{a^n b^n \mid n \geq 1\} \cdot \{c\}^+ \\ \{a^m b^n c^n \mid n \geq 1, m \geq 1\} &= \{a\}^+ \cdot \{b^n c^n \mid n \geq 1\}\end{aligned}$$

sind in $\mathcal{L}_2$:

- $\{a^n b^n \mid n \geq 1\} \in \mathcal{L}_2$ und $\{b^n c^n \mid n \geq 1\} \in \mathcal{L}_2$
- $\{a\}^+ \in \mathcal{L}_3 \subseteq \mathcal{L}_2$ und $\{c\}^+ \in \mathcal{L}_3 \subseteq \mathcal{L}_2$
- $\mathcal{L}_2$ ist unter Konkatenation abgeschlossen.

Dresden

(2) $\{a^nb^nc^n \mid n \geq 1\} = \{a^nb^nc^m \mid n,m \geq 1\} \cap \{a^mb^nc^n \mid n,m \geq 1\}$.

Wäre $\mathcal{L}_2$ unter $\cap$ abgeschlossen, so würde $\{a^nb^nc^n \mid n \geq 1\} \in \mathcal{L}_2$ folgen.

(3) $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.

Wäre $\mathcal{L}_2$ unter Komplement abgeschlossen, so auch unter $\cap$.

Beachte:

Man kann daher das Äquivalenzproblem für kontextfreie Sprachen nicht einfach auf das Leerheitsproblem reduzieren.

In der Tat kann man zeigen, daß das Äquivalenzproblem für kontextfreie Sprachen sogar unentscheidbar ist.

Dresden