Universität Siegen
Lehrstuhl Theoretische Informatik
Markus Lohrey

Übungsblatt 10

Aufgabe 1. Sei $G = (\{S, A, B\}, \{a, b\}, P, S)$ eine kontextfreie Grammatik, wobei P gegeben ist durch

$$\begin{aligned} S &\to ASB \mid ab \\ A &\to aAS \mid a \\ B &\to SbS \mid A \mid bb. \end{aligned}$$

Geben Sie eine Grammatik G' in CNF so an, dass $L(G') = L(G)$.

Lösung zu Aufgabe 1. Wir verwenden das Verfahren zur Umwandlung einer kontextfreien Grammatik in CNF wie ab Folie 196 im Skript beschrieben.

Schritt 1

G enthält keine Produktion der Form $S \to \varepsilon$.

Schritt 2

Einführen neuer Variablen A_a mit Produktionen $A_a \to a$ für jedes Terminalsymbol $a \in \Sigma$ und Ersetzen aller Vorkommen von a in einer rechten Seite, die nicht bereits nur aus einem Alphabetsymbol auf der rechten Seite bestehen. Änderungen an den Produktionen sind **fettgedruckt** markiert.

- $S \to ASB \mid \boldsymbol{A_a A_b}$
- $A \to \boldsymbol{A_a} AS \mid a$
- $B \to S\boldsymbol{A_b}S \mid A \mid \boldsymbol{A_b A_b}$
- $\boldsymbol{A_a \to a}$
- $\boldsymbol{A_b \to b}$

Schritt 3

Elimination von Kettenregeln der Form $X \to Y$ durch Ersetzen von Y durch die rechten Seiten der Produktion von Y.

- $S \to ASB \mid A_aA_b$
- $A \to A_aAS \mid a$
- $B \to SA_bS \mid \boldsymbol{A_aAS} \mid \boldsymbol{a} \mid A_bA_b$
- $A_a \to a$
- $A_b \to b$

Bemerkung: Falls es Zyklen gibt (wie $A \to B$ und $B \to A$), wird man irgendwann bei $A \to A$ angelangen. Diese Produktion muss man einfach weglassen.

Schritt 4

Elimination von Produktionen der Form $X \to X_1 \dots X_n$ mit $n \geq 3$ durch Einführen neuer Nicht-Terminale. Beispielsweise schreiben wir $S \to ASB$ um in $S \to AB_1$ und fügen ein neues Nicht-Terminal B_1 hinzu mit einer Produktion $B_1 \to SB$.

- $S \to A\boldsymbol{B_1} \mid A_aA_b$
- $A \to A_a\boldsymbol{B_2} \mid a$
- $B \to S\boldsymbol{B_3} \mid A_a\boldsymbol{B_2} \mid a \mid A_bA_b$
- $A_a \to a$
- $A_b \to b$
- $\boldsymbol{B_1 \to SB}$
- $\boldsymbol{B_2 \to AS}$
- $\boldsymbol{B_3 \to A_bS}$

Aufgabe 2. Für ein Wort $w = a_1 \cdots a_n \in \Sigma^*$ (mit $a_1, \dots, a_n \in \Sigma$) ist das Spiegelwort w^r definiert als $w^r = a_n \cdots a_1$. Beweisen Sie, dass die folgenden Sprachen über $\Sigma = \{a, b, c\}$ kontextfrei sind.

(a) $L' = \{vcwcw^r \mid v, w \in \{a, b\}^*\}$

(b) $L'' = \{wcw^rcv \mid v, w \in \{a, b\}^*\}$

Lösung zu Aufgabe 2.

(a) Es gibt eine kontextfreie Grammatik G', die diese Sprache erzeugt.

- $G' = (\{S, A, B\}, \{a, b, c\}, P, S)$
- $P = \{S \to AcB \mid cB,\ A \to a \mid b \mid aA \mid bA,\ B \to aBa \mid bBb \mid c\}$

Dabei erzeugt A beliebige Wörter über $\{a, b\}$ (mindestens Länge 1) und B erzeugt wcw^r für $w \in \{a, b\}^*$.

(b) Es gibt eine kontextfreie Grammatik G'', die diese Sprache erzeugt.

- $G'' = (\{S, A, B\}, \{a, b, c\}, P, S)$
- $P = \{S \to AcB \mid Ac,\ A \to aAa \mid bAb \mid c,\ B \to a \mid b \mid aB \mid bB\}$

Dabei erzeugt A hier ein Wort wcw^r und B erzeugt eine beliebiges Wort über $\{a, b\}$ der Länge mindestens 1.

Aufgabe 3. Zeigen Sie mit Hilfe des Pumping-Lemmas für kontextfreie Sprachen, dass die folgenden Sprachen nicht kontextfrei sind.

(a) $L_1 = \{a^{n^2} \mid n \geq 0\}$

(b) $L_2 = \{ww \mid w \in \{a, b\}^*\}$

(c) $L_3 = L' \cap L''$ (mit L', L'' wie in Aufgabe 2)

Lösung zu Aufgabe 3.

(a) Sei $n \in \mathbb{N}$ beliebig. Wähle $z = a^{n^2} \in L_1$. Es gilt $|z| = n^2 \geq n$.

Betrachte alle Zerlegungen $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.

Wir haben $u = a^b, v = a^c, w = a^d, x = a^e, y = a^f$ $(b+c+d+e+f = n^2)$. Zudem gilt $c + e \geq 1$ und $c + d + e \leq n$.

Wir wählen den Pumpfaktor $i = 2$ und betrachten $uv^i wx^i y$:

$uv^2wx^2y = a^{b+2c+d+2e+f} = a^{n^2+c+e}$

Nun müssen wir zeigen, dass $n^2 + c + e$ keine Quadratzahl ist und somit $uv^2wx^2y \notin L_1$.

Es gilt $n^2 < n^2 + c + e < (n + 1)^2$ (vergleiche Übung 7, Aufgabe 3a).

$n^2 < n^2 + c + e$ gilt, da $c + e \geq 1$. Außerdem gilt

$$
\begin{aligned}
& n^2 + c + e \\
\leq & n^2 + c + d + e \\
\leq & n^2 + n && \text{(da } c + d + e \leq n\text{)} \\
< & n^2 + 2n + 1 \\
= & (n + 1)^2
\end{aligned}
$$

Also gilt $uv^2wx^2y \notin L_1$ und somit ist die Sprache ist nicht kontextfrei.

(b) Sei $n \in \mathbb{N}$ beliebig. Wähle $z = a^n b^n a^n b^n \in L_2$. Es gilt $|z| = 4n \geq n$.

Betrachte alle Zerlegungen $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.

Wir unterscheiden drei mögliche Positionen an denen sich vwx im Wort z befinden kann:

1. Vollständig in der ersten Hälfte, $\boldsymbol{a^n b^n} a^n b^n$
2. Vollständig in der zweiten Hälfte, $a^n b^n \boldsymbol{a^n b^n}$
3. In der Mitte, $a^n \boldsymbol{b^n a^n} b^n$

Da $|vwx| \leq n$ sind so alle Möglichkeiten abgedeckt.

Fall 1, vwx liegt in der ersten Hälfte ($\boldsymbol{a^n b^n} a^n b^n$):

Wir wählen den Pumpfaktor $i = 0$ und betrachten $uv^i wx^i y$:

Wir erhalten ein Wort $uv^0 wx^0 y = uwy = a^{n-k} b^{n-j} a^n b^n$ mit $k \geq 0$ und $j \geq 0$.

Sei $l = k + j$. Es gilt $l \geq 1$ (da $|vx| \geq 1$) und $l \leq n$ (da $|vwx| \leq n$).

Falls l ungerade ist, kann uwy nicht in zwei gleiche Wörter zerlegt werden und liegt nicht in der Sprache.

Falls l gerade ist, lässt sich uwy in zwei Wörter gleicher Länge w_1, w_2 zerlegen, wobei $w_1 = a^{n-k} b^{n-j} a^p$ und $w_2 = a^{n-p} b^n$ mit $p = \frac{l}{2}$: Es ist

$$\begin{aligned} |w_1| =& (n-k) + (n-j) + p \\ =& 2n - 2p + p \\ =& (n-p) + n = |w_2|. \end{aligned}$$

Da $p \geq 1$ (wegen $l \geq 1$), endet w_1 mit a, w_2 hingegen mit b und somit $uwy \notin L_2$.

Beispiel: Sei $z = a^4 b^4 a^4 b^4$ und $vx = ab, w = \varepsilon$. Nach dem Aufpumpen erhalten wir $uwy = a^3 b^3 a^4 b^4$, $w_1 = a^3 b^3 a$ und $w_2 = a^3 b^4$.

Fall 2, vwx liegt in der zweiten Hälfte ($a^n b^n \boldsymbol{a^n b^n}$)

Analog zu Fall 1.

Für den Pumpfaktor $i = 0$ hat uwy entweder ungerade Länge, oder, falls uwy sich in zwei Wörter gleicher Länge w_1, w_2 zerlegen lässt, so haben diese die Form $w_1 = a^n b^{n-p}$ und $w_2 = b^p a^{n-i} b^{n-j}$.

Da $p \geq 1$ beginnt w_1 mit einem a, w_2 hingegen mit einem b. Somit ist auch hier $uwy \notin L_2$.

Fall 3, vwx in der Mitte ($a^n \boldsymbol{b^n a^n} b^n$)

Wir wählen wieder den Pumpfaktor $i = 0$.

Das Wort uwy hat die Form $a^n b^{n-k} a^{n-j} b^n$, wobei $k \geq 0$ und $j \geq 0$ und $k + j \geq 1$ sowie $k + j \leq n$.

Wenn $k + j$ ungerade ist, dann hat auch uwy ungerade Länge und kann somit nicht in der Sprache liegen.

Andernfalls betrachten wir nun die möglichen Zerlegungen in Wörter w_1, w_2 gleicher Länge:

- Fall $k = j$: $w_1 = a^n b^{n-k}$, $w_2 = a^{n-j} b^n$
- Fall $j < k$: $w_1 = a^n b^{n-k} a^q$, $w_2 = a^{n-j-q} b^n$ mit $q = \frac{k-j}{2} > 0$
- Fall $j > k$: $w_1 = a^n b^{n-k-q}$, $w_2 = b^q a^{n-j} b^n$ mit $q = \frac{j-k}{2} > 0$

In jedem Fall ist $w_1 \neq w_2$ und folglich $uwy \notin L_2$. Damit ist L_2 nicht kontextfrei.

(c) $L_3 = L' \cap L'' = \{scs^r cs \mid s \in \{a, b\}^*\}$. Sei $n \in \mathbb{N}$ beliebig. Wähle $z = a^n ca^n ca^n \in L_3$. Es gilt $|z| = 3n + 2 \geq n$.

Betrachte alle Zerlegungen $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.

Fall 1, c ist in vx enthalten

Wir wählen den Pumpfaktor $i = 2$ und betrachten $uv^i wx^i y$:

$uv^2 wx^2 y$ enthält dann mindestens 3 c's, aber jedes Wort in L_3 enthält genau 2 c's. Also gilt $uv^2 wx^2 y \notin L_3$.

Fall 2, c ist nicht in vx enthalten

Das Wort vx besteht also nur aus a's.

Wir wählen den Pumpfaktor $i = 2$ und betrachten $uv^i wx^i y$:

Falls w ein c enthält, erhalten wir ein Wort $a^k ca^j ca^n$ (bzw. $a^n ca^k ca^j$) mit $k > n$ oder $j > n$, und somit $uv^2 wx^2 y \notin L_3$. Falls $w \in L(a^*)$ erhalten wir ein Wort $a^k ca^n ca^n$ (bzw. $a^n ca^k ca^n$ oder $a^n ca^n ca^k$) mit $k > n$ und somit gilt wiederum $uv^2 wx^2 y \notin L_3$.

Somit ist die Sprache L_3 nicht kontextfrei.

Aufgabe 4. Zeigen Sie die folgenden Aussagen:

(a) Sei L eine kontextfreie Sprache über Σ und sei $h : \Sigma^* \to \Gamma^*$ ein Homomorphismus. Dann ist $h(L)$ wieder kontextfrei.

(b) $L = \{baba^2ba^3b \cdots ba^{n-1}ba^nb \mid n \geq 1\}$ ist nicht kontextfrei.

Hinweis: Nutzen Sie die in (a) gezeigte Abschlusseigenschaft, zusammen mit Aufgabe 3 Teil (a).

Lösung zu Aufgabe 4.

(a) Sei $G = (V, \Sigma, P, S)$ eine Grammatik mit $L(G) = L$. Man erhält eine kontextfreie Grammatik zu $h(L)$, indem man bei den rechten Seiten aller Produktionen jedes Terminalsymbol $a \in \Sigma$ durch $h(a)$ ersetzt. Am einfachsten sieht man das, wenn man annimmt, dass G in Chomsky-Normalform ist. Denn dann kann man G' wie folgt definieren: $G' = (V, \Gamma, P', S)$ mit

$$P' = \{A \to BC \mid (A \to BC) \in P\} \cup \{A \to h(a) \mid (A \to a) \in P\}.$$

Es gilt $h(L) = L(G')$, da man aus S die gleichen Strings über V ableiten kann. Zu gegebenem $\alpha \in V^*$ leitet man mit P dann w und mit P' das Wort $h(w)$ ab (Homomorphismeneigenschaft: $h(w) = h(a_1 \cdots a_n) = h(a_1) \cdots h(a_n)$). Die neue Grammatik G' muss dabei nicht wieder in CNF sein, aber sie ist offensichtlich kontextfrei.

(b) Sei $h : \{a, b\}^* \to \{a\}^*$ definiert durch $h(b) = a$ und $h(a) = aa$. Dann besteht $h(L)$ nur aus Wörtern über $\{a\}^*$ der Länge

$$1 + 2 + 1 + 4 + \cdots + 1 + 2(n-1) + 1 + 2n + 1 = (n+1) + 2\sum_{k=1}^{n} k.$$

Der letzte Teil ist die Gaußsumme und wertet sich zu $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ aus. Also kann man $h(L)$ wie folgt beschreiben:

$$h(L) = \{a^{(n+1)^2} \mid n \geq 1\}$$

Wäre L kontextfrei, so müsste auch $h(L)$ kontextfrei sein (h ist ein Homomorphismus). Allerdings ist $h(L)$ bis auf die zwei Elemente $\varepsilon = a^0$ und a^1 genau gleich L_1 aus Aufgabe 3, wo wir bereits mit Hilfe des Pumping-Lemmas gezeigt haben, dass diese nicht kontextfrei ist. Somit kann L nicht kontextfrei sein.