# Theoretische Informatik
# 2. Übungsblatt

Paul Winkler
11818749

**Aufgabe 1.** Die Sprache $L = \{wc^n \mid w \in \{a,b\}^*, n \in \mathbb{N}, n_a(w) = n \text{ oder } n_b(w) = n\}$ wird von folgender Grammatik $G$ erzeugt: $G = \langle \{S, A, B\}, \{a, b\}, P, S\rangle$, wobei

$$P = \begin{cases} S \to A \mid B \\ A \to aAc \mid bA \mid \varepsilon \\ B \to bBc \mid aB \mid \varepsilon. \end{cases}$$

Das sieht man leicht, indem man $L$ aufspaltet in $\{wc^n \mid n_a(w) = n\} \cup \{wc^n \mid n_b(w) = n\}$, für beide Teile die offensichtliche Grammatik nimmt und die beiden Grammatiken dann vereinigt.

**Aufgabe 2.** Die Sprache $L = \{ww \mid w \in \{a,b\}^*\}$ ist nicht kontextfrei. Wir führen einen Widerspruchsbeweis: Angenommen, $L$ wäre kontextfrei. Nach dem Schleifensatz gibt es dann ein $n \in \mathbb{N}$, sodass alle $w \in L$ mit $|w| \geqslant n$ eine Darstellung der Form $w = v_1v_2v_3v_4v_5$ haben, wobei

$$(i)\ \ v_2v_4 \neq \varepsilon, \qquad (ii)\ \ |v_2v_3v_4| \leqslant n, \qquad (iii)\ \ \forall k \geqslant 0\colon v_1v_2^kv_3v_4^kv_5 \in L.$$

Sei $w = a^nb^na^nb^n \in L$ und sei $w = v_1v_2v_3v_4v_5$ eine solche Darstellung. Wir machen eine Fallunterscheidung, je nachdem, wo der mittlere Teil $v_2v_3v_4$ steht, und zeigen, dass $v_1v_3v_5 \notin L$:

1. $v_2v_3v_4$ liegt in der ersten Hälfte von $w$:

   Dann gibt es natürliche Zahlen $k, l$ sodass $v_1v_3v_5 = a^{n-k}b^{n-l}a^nb^n$. Wegen $(i)$ und $(ii)$ gilt $\{k, l\} \neq \{0\}$ und $k + l \leqslant n$; daher beginnt die erste Worthälfte von $v_1v_3v_5$ mit $a$, die zweite aber mit $b$, Widerspruch.

2. $v_2v_3v_4$ liegt in der Mitte von $w$:

   2.1. $v_2 = b^ma^l$ und $v_4 = a^k$:
   Dann gilt $v_1v_3v_5 = a^nb^{n-m}a^{n-l-k}b^n$. Falls $m > l + k$, dann endet die erste Worthälfte mit $a$, die zweite aber mit $b$, Widerspruch. Ist hingegen $m < l + k$, dann beginnt die erste Worthälfte mit $a$, die zweite aber mit $b$, Widerspruch. Im Fall $m = l + k$ ist $a^nb^{n-m}$ die erste, aber $a^{n-m}b^n$ die zweite Worthälfte, Widerspruch.

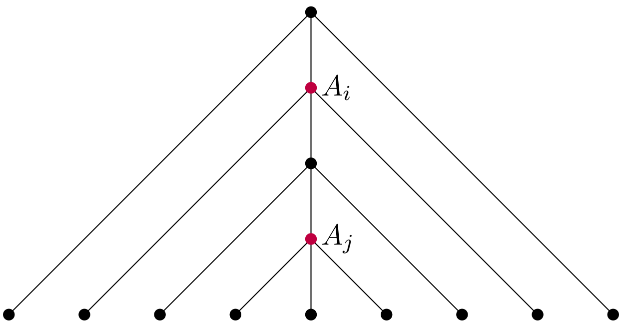
   2.2. $v_2 = b^m$ und $v_4 = b^k$:
   Dann ist $v_1v_3v_5 = a^nb^{n-k-m}a^nb^n$. Wegen $k + m \neq 0$ endet die erste Worthälfte auf $a$, die zweite aber auf $b$, Widerspruch.

Die anderen Fälle sind symmetrisch dazu.

**Aufgabe 3.** Wir wissen bereits, dass eine Grammatik $G$ existiert, die linear, frei von Umbenennungen und ohne $\varepsilon$-Produktionen ist und $L$ erzeugt. Sei

$$n = \max\{|\alpha| \colon S \Longrightarrow_G^{\leqslant |N|+1} \alpha, \alpha \in (N \cup T)^*\} + 1.$$

Sei $w \in L$ mit $|w| \geqslant n$, dann lässt sich $w$ nicht in $|N|+1$ Schritten herleiten und hat damit einen Ableitungsbaum $B$ mit Höhe $h > |N|+1$. (Hier geht ein, dass $L$ linear ist, denn wegen der speziellen Form des Ableitungsbaums können wir aus der Anzahl der benötigten Schritte auf dessen Höhe schließen – sie sind nämlich ident.) Seien $A_1, \ldots, A_{|N|+1}$ die ersten $|N|+1$ Nichtterminalsymbole auf dem längsten Pfad von $B$. Nach dem Schubfachprinzip existieren $i < j \leqslant |N|+1$ und ein Nichtterminal $A$ mit $A_i = A = A_j$.




Die Ableitung von $w$ hat die Form

$$S \Longrightarrow_G^* v_1 A v_5 \Longrightarrow_G^* v_1 v_2 A v_4 v_5 \Longrightarrow_G^* v_1 v_2 v_3 v_4 v_5.$$

Wegen $A_i = A_j$ lässt sich somit auch $v_1 v_2^k v_3 v_4^k v_5$ für alle $k \geqslant 0$ ableiten. Das Wort $v_1 v_2 A v_4 v_5$ lässt sich mit einem Ableitungsbaum der Höhe $j \leqslant |N|+1$, also in höchstens $|N|+1$ Schritten ableiten, daher gilt $|v_1 v_2 v_4 v_5| \leqslant |v_1 v_2 A v_4 v_5| \leqslant n$. Weil $G$ frei von Umbenennungen und ohne $\varepsilon$-Produktionen ist, gilt $|v_2 v_4| \neq \varepsilon$.

**Aufgabe 4.**

- regulär $\subset$ linear:

  Wir wissen bereits, dass eine Sprache genau dann regulär ist, wenn sie von einer strikt rechtslinearen Grammatik erzeugt wird. Offenbar ist jede strikt rechtslineare Grammatik auch linear, also gilt die Inklusion. Für die Striktheit betrachte die Sprache $L = \{a^i b^i \mid i \in \mathbb{N}\}$, die bekannterweise nicht regulär ist. Sie wird von der linearen Grammatik $G = \langle \{S\}, \{a, b\}, P, \{S\} \rangle$ mit $P = \{S \to aSb \mid \varepsilon\}$ erzeugt und ist somit linear.

- linear $\subset$ kontextfrei:

  Das überhaupt die richtige Teilmengenbeziehung besteht, ist nach Definition klar. Betrachte die Sprache $L = \{a^i b^i c^j d^j \mid i, j \in \mathbb{N}\}$. Die Sprache $L_1 = \{a^i b^i \mid i \in \mathbb{N}\}$ ist wie oben gezeigt kontextfrei. Natürlich ist auch $L_2 = \{c^j d^j \mid j \in \mathbb{N}\}$ kontextfrei und damit auch $L = L_1 \cdot L_2$.

  Um zu zeigen, dass $L$ nicht linear ist, verwenden wir den Schleifensatz aus Aufgabe 3. Angenommen, $L$ wäre linear. Seien $n \in \mathbb{N}$ und $w = v_1 v_2 v_3 v_4 v_5 = a^i b^i c^j d^j$ wie dort gewählt, wobei wir zusätzlich $i, j > n$ fordern. Weil nun $|v_1 v_2 v_4 v_5| \leqslant n$ ist, muss $v_2 = a^k$ sowie $v_4 = d^l$ für geeignete $k, l \in \mathbb{N}$ sein. Betrachten wir nun das Wort $w' = v_1 v_3 v_5 = a^{i-k} b^i c^j d^{j-l}$: es liegt wegen $v_2 v_4 \neq \varepsilon$, d. h. $k \neq 0$ oder $l \neq 0$, nicht mehr in $L$, Widerspruch.

**Aufgabe 5.** $G$ ist mehrdeutig, weil das Wort $a + b \cdot c$ auf zwei verschiedene Weisen abgeleitet werden kann:


Die beiden Bäume sind offenbar ungleich. $\mathrm{L}(G)$ ist aber nicht inhährent mehrdeutig, denn folgende eindeutige Grammatik erzeugt $\mathrm{L}(G)$: $G' = \langle \{S', A, M, V\}, \{a, b, c, +, \cdot\}, P', S' \rangle$, wobei

$$P' = \begin{cases} S' & \to A \mid M \\ A & \to A + A \mid M \\ M & \to M \cdot M \mid V \\ V & \to a \mid b \mid c. \end{cases}$$

Die Grammatik ist eindeutig, weil wir festlegen, dass zuerst alle Additionen erzeugt werden müssen und erst danach allfällige Multiplikationen (also gleichsam Punkt vor Strich).

**Aufgabe 6.** Sei $L$ eine reguläre Sprache und $D = \langle Q, A, \delta, q_0, F \rangle$ ein DFA mit $\mathrm{L}(D) = \{w \in A^* \mid \delta(q_0, w) \in F\} = L$. Wir definieren die Grammatik $G = \langle Q, A, P, q_0 \rangle$, wobei

$$P = \{q \to a\delta(q, a) \mid q \in Q, a \in A\} \cup \{q \to \varepsilon \mid q \in F\}.$$

Zuerst zeigen wir $\mathrm{L}(D) = \mathrm{L}(G)$:

»$\subseteq$« Sei $w = x_1 \cdots x_n \in \mathrm{L}(D)$ beliebig, dann gibt es Zustände $q_1, \ldots, q_n$ mit $q_n \in F$, sodass

$$\delta(q_{i-1}, x_i) = q_i, \quad i = 1, \ldots, n.$$

Damit existiert die Ableitung

$$q_0 \implies x_1 q_1 \implies x_1 x_2 q_2 \implies \cdots \implies x_1 \cdots x_n q_n \implies x_1 \cdots x_n.$$

»$\supseteq$« Sei $w = x_1 \cdots x_n \in \mathrm{L}(G)$ beliebig, dann gibt es eine Ableitung $q_0 \implies x_1 q_1 \implies x_1 x_2 q_2 \implies \cdots \implies x_1 \cdots x_n q_n \implies x_1 \cdots x_n$, in der die Produktionen $q_0 \to x_1 q_1, \ldots, q_{n-1} \to x_n q_n, q_n \to \varepsilon$ verwendet werden, d. h. $q_n \in F$. Für $0 \leqslant i < j \leqslant n$ gilt $\delta(q_i, x_{i+1} \cdots x_j) = q_j$, ergo $\delta(q_0, w) = q_n$ und somit $w \in \mathrm{L}(N)$.

Es bleibt zu zeigen, dass $G$ eindeutig ist. Angenommen, es gäbe ein Wort $w \in L$ mit zwei verschiedenen Ableitungen. Dann gäbe es eine erste Stelle, an der sich die Ableitungen unterscheiden, d. h. ein $i \leqslant n$ mit

$$\begin{aligned} x_1 \cdots x_i q_i &\implies x_1 \cdots x_i x_{i+1} q_{i+1} \\ x_1 \cdots x_i q_i &\implies x_1 \cdots x_i x_{i+1} q'_{i+1} \end{aligned}$$

und $q_{i+1} \neq q'_{i+1}$. Dann enthält $G$ die beiden Produktionsregeln

$$\begin{aligned} q_i &\to x_{i+1} q_{i+1} \\ q_i &\to x_{i+1} q'_{i+1}, \end{aligned}$$

was nach Definition von $G$ $\delta(q_i, x_{i+1}) = q_{i+1}$ und $\delta(q_i, x_{i+1}) = q'_{i+1}$ impliziert – Widerspruch.