Theoretische Informatik

2. Übung

Richard Weiss

25.4.2021

Aufgabe 1. Sei $L = \{wc^n \mid w \in \{a,b\}^*, n_a(w) = n \text{ oder } n_b(w) = n\}$. Geben Sie eine kontextfreie Grammatik an die L erzeugt.

Lösung. Die fehlende Quantisierung für n lässt 2 zulässige Interpretationen von L zu. Gemeint war die Zeile

1. Interpretation ($L = L_n := \{wc^n \mid w \in \{a,b\}^*, n_a(w) = n \text{ oder } n_b(w) = n\}$, $n \in \mathbb{N}$):

   An diese habe ich ursprünglich gedacht.

   Zwecks besseren Verständnisses, listen wir die Elemente von L für verschiedene $n \in \mathbb{N}$ auf. Dabei bezeichnet $x \cdot x = x^m$ für irgendein beliebiges $m \in \mathbb{N}$. Wenn der Ausdruck mehrmals vorkommt, dürfen die m auch unterschiedlich sein.

   - $n = 0$:

   $$a \cdots a$$
   $$b \cdots b$$

   - $n = 1$:

   $$(a \cdots a)b(a \cdots a)c$$
   $$(b \cdots b)a(b \cdots b)c$$

   - $n = 2$:

   $$(a \cdots a)b(a \cdots a)b(a \cdots a)c^2$$
   $$(b \cdots b)a(b \cdots b)a(b \cdots b)c^2$$

   - $n \in \mathbb{N}$:

   $$\underbrace{((a \cdots a)b) \cdots ((a \cdots a)b)}_{n\text{-mal}}(a \cdots a)c^n$$
   $$\underbrace{((b \cdots b)a) \cdots ((b \cdots b)a)}_{n\text{-mal}}(b \cdots b)c^n$$

Wir wollen nun all diese Wörter systematisch aufbauen, mit einem endlichem Regelwerk P. Sei dabei aber o.B.d.A. $n \neq 0$.

$$P = \begin{cases} S \to X^n C^n \mid Y^n C^n, \\ X \to XB \mid BX, \\ Y \to YA \mid AY, \\ A \to a, \\ B \to b, \\ C \to c \end{cases}$$

Bsp: $n=1$, $abac \in L_1$ kann aber nicht abgeleitet werden (X leitet nur B's ab, Y nur A's)

besser P_n = $S \to (Ab)^n A c^n$
$S \to (Ba)^n B c^n$
$A \to aA \mid \varepsilon$
$B \to bB \mid \varepsilon$

Die kontextfreie Grammatik $G = \langle N, T, P, S \rangle$ erzeugt L.

$$N = \{S, X, Y, A, B, C\}, \quad T = \{a, b, c\}$$

2. Interpretation ($L = \bigcup_{n \in \mathbb{N}} L_n$):

   Man wäre möglicherweise nun dazu geneigt, Satz 2.2 zu verwenden; d.h. die Abschlusseigenschaft kontextfreier Sprachen unter Vereinigung. Diese wurde aber nur für endliche Vereinigungen bewiesen. Wir gehen also zu Fuß mit $G = \langle N, T, P, S \rangle$, wobei

$$N = \{S, A, B\}, \quad T = \{a, b\}, \quad P = \begin{cases} S \to A \mid B, \\ A \to aAc \mid bA \mid \varepsilon, \\ B \to bBc \mid aB \mid \varepsilon. \end{cases}$$

Antwort richtig, 1. Interpretation „außer Konkurrenz"

Die missverständliche Formulierung der Angabe kostet zwar etwas Zeit, wirft aber gleichzeitig die Frage auf, wann unendliche Vereinigeungen (oder vielleicht sogar Grenzwerte) von kontextfreien Sprachen wieder kontextfrei sind.

□

Aufgabe 2. Ist die Sprache $L = \{ww \mid w \in \{a, b\}^*\}$ kontextfrei? Begründen Sie Ihre Antwort.

Lösung. L ist nicht kontextfrei. Ansonsten sei n wie im Lemma 2.7 (Schleifensatz (pumping lemma) für kontextfrei Sprachen). $w := a^n b^n a^n b^n \in L$ kann geschrieben werden als $v_1 v_2 v_3 v_4 v_5$, wobei

1. $v_2 v_4 \neq \varepsilon$,
2. $|v_2 v_2 v_3| \leq n$, und
3. für alle $k \geq 0$ ist auch $w_k := v_1 v_2^k v_3 v_4^k v_5 \in L$.

Wegen 2., gibt es ein $i = 0, \ldots, n$, sodass $v_2 v_3 v_4$ ein Teil-Wort von $a^i b^{n-i}$ oder $b^i a^{n-i}$ ist. Wegen 1., ist $v_2 \neq \varepsilon$ oder $v_4 \neq \varepsilon$. Auf jeden Fall kann man w so wie in 3. „aufpumpen" und aus L rausfliegen. Man führt nämlich zu viele a-s oder b-s in der lokalen a- bzw. b-Sequenz ein, oder a-s und / oder b-s in angrenzenden Sequenzen (pro Sequenz aber jeweils nur a-s oder b-s). Widerspruch!

□

Aufgabe 3. Eine kontextfreie Grammatik $G = \langle N, T, P, S \rangle$ heißt *linear* wenn jede Produktion von der Form $A \to uBv$ oder $A \to u$ ist wobei $B \in N$, $u, v \in T^*$. Eine Sprache L heißt linear falls eine lineare Grammatik G existiert, mit $\mathrm{L}(G) = L$. Beweisen Sie den folgenden Schleifensatz (pumping lemma) für lineare Sprachen:

Satz. Sei L eine lineare Sprache. Dann gibt es ein $n \in \mathbb{N}$ so dass jedes $w \in L$ mit $|w| \geq n$ geschrieben werden kann als $w = v_1 v_2 v_3 v_4 v_5$ so dass

1. $v_2v_4 \neq \varepsilon$,
2. $|v_1v_2v_4v_5| \leq n$, und
3. für alle $k \geq 0$ ist auch $v_1v_2^kv_3v_4^kv_5 \in L$.

Hinweis: Entfernen Sie zunächst alle Umbenennungen aus G. Setzen Sie $n = m_{|N|+1}$ *wobei* $m_k = \max\left\{|u| + |v| \mid S \Longrightarrow_G^{\leq k} uAv\right\}$.

Lösung. Sei $G = \langle N, T, P, S\rangle$ eine lineare Grammatik mit $\mathrm{L}(G) = L$. O.B.d.A, habe G (bzw. P) keine Umbenennungen. (Zur Entfernung gehe analog wie im Beweis von Lemma 2.4 vor.) O.B.d.A, sei $T \neq \emptyset$. (Sonst wähle $n \geq 1$; dann gibt es keine $w \in L$ mit $|w| \geq n$.)
Setze $n = m_{|N|+1}$ wobei $m_k = \max\left\{|u| + |v| \mid \exists\, A \in N : S \Longrightarrow_G^{\leq k} uAv\right\}$.
Offenbar gilt $m_{k_1} \leq m_{k_2}$ für $k_1 \leq k_2$. Weil $T \neq \emptyset$, gilt über dies $m_{k+1} \geq m_k + 2$. Dazu erweitert man einfach die maximierende Ableitung von m_k durch ein $w \in T$.

!?

$$S \Longrightarrow_G^{\leq k} uAv \Longrightarrow_G uwAwv, \quad S \Longrightarrow_G^{\leq k+1} uwAwv$$

B ?! wieso ist B → wAw in P ?
bzw A → wAw ?

Somit bekommt man die Abschätzung

$$m_{k+1} \geq |uw| + |vw| = |u| + |v| + 2|w| = |u| + |v| + 2 \geq m_k + 2.$$

Sei $w \in L$ mit $|w| \geq n$.
Angenommen, w hat eine Ableitung, sodass $A_1, \ldots, A_k$, $k \leq |N|$ alle Nichtterminalsymbole sind, d.h.

$$S \Longrightarrow_G l_1A_1r_1 \Longrightarrow_G \cdots \Longrightarrow_G l_1 \cdots l_kA_kr_k \cdots r_1 \Longrightarrow_G w,$$

oder

$$S \Longrightarrow_G^k l_kA_kr_k \cdots r_1 \Longrightarrow_G w.$$

Damit erhalten wir aber den Widerspruch

$$m_{|N|+1} - 1 \geq m_{k+1} - 1 \geq m_k + 1 \geq |l_k| + |r_k| + 1 \geq |w| \geq n = m_{|N|+1} > m_{|N|+1} - 1.$$

Jede Ableitung von w enthält also zwangsläufig $A_1, \ldots, A_{|N|+1} \in N$. Seien diese die ersten $|N| + 1$. Nach dem Schubfachprinzip, gibt es nun $i, j = 1, \ldots, |N|$, mit $i < j$, sodass $A_i = A_j =: A$. Eine Ableitung kann man nun schreiben als

$$S \Longrightarrow_G^* v_1Av_5 \Longrightarrow_G^* v_1v_2Av_4v_5 \Longrightarrow_G^* w,$$

oder

$$S \Longrightarrow_G^{\leq |N|+1} v_1v_2Av_4v_5 \Longrightarrow_G^* w.$$

1. gilt, weil G keine Umbenennungen hat und wegen der ersten oberen Darstellung. Aus dieser liest man auch 3. heraus, indem man den zweiten $\Longrightarrow_G^*$ einfach k-mal widerholt. 2. folgt aus der zweiten oberen Darstellung und der Definition von $m_{|N|+1} = n$, weil

$$|v_1v_2v_4v_5| = |v_1v_2| + |v_4v_5| \in \left\{|u| + |v| \mid \exists\, A \in N : S \Longrightarrow_G^{\leq |N|+1} uAv\right\}.$$

$\square$

Aufgabe 4. Zeigen Sie dass die regulären Sprachen strikt in den linearen Sprachen enthalten sind und die linearen Sprachens trikt in den kontextfreien.

Hinweis: $\{a^i b^i c^j d^j \mid i, j \in \mathbb{N}\}$.

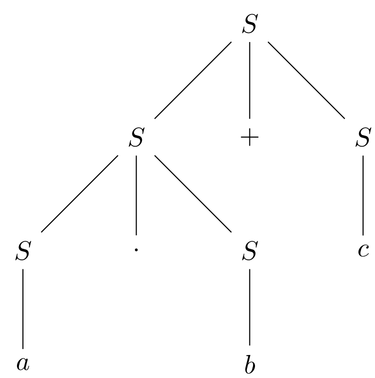
Lösung. Die erste Behauptung folgt aus Satz 2.1 und Beispiel 2.5.
Betrachte nun die Sprache $L := \{a^i b^i c^j d^j \mid i, j \in \mathbb{N}\}$. Diese ist, wegen Satz 2.2, als Quadrat der Sprache aus Beispiel 2.5, kontextfrei. Angenommen, sie wäre linear. Sei n so wie im Satz (Schleifensatz (pumping lemma) für lineare Sprachen) der Aufgabe 3.
Das Wort $a^n b^n c^n d^n \in L$ lässt sich nun leicht aus L „rauspumpen". Wegen der 1. und 3. Bedingung, kann man überhaupt „pumpen"; und wegen der 2. passiert das nur lokal in einer, oder zwei angrenzenden Sequenzen desselben Symbols.
Wir haben dabei übrigens auch gezeigt, dass das Produkt zweier linearen Sprachen im Allgemeinen nicht linear ist. □

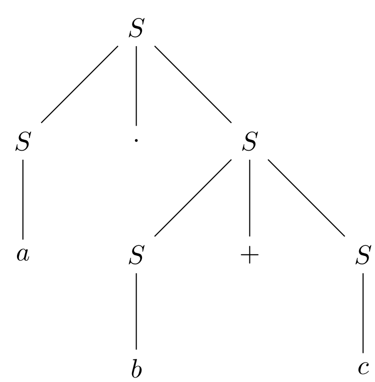
Aufgabe 5. Sei $G = \langle\{S\}, \{a, b, c, +, \cdot\}, P, S\rangle$ wobei $P =$

$$S \to S + S \mid S \cdot S \mid a \mid b \mid c.$$

Zeigen Sie dass G mehrdeutig ist. Ist $\mathrm{L}(G)$ inhärent mehrdeutig?

Lösung. Betrachte die beiden Ableitungsbäume für $a \cdot b + c$.

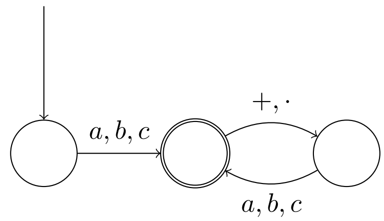




Offenbar gilt

$$\mathrm{L}(G) = L := \left\{ \left(\prod_{i=1}^{n} x_i y_i \right) x_{n+1} \mid n \in \mathbb{N},\ x_1, \ldots, x_{n+1} \in \{a, b, c\},\ y_1, \ldots, y_n \in \{+, \cdot\} \right\}.$$

Folgender NFA zeigt, dass L regulär ist.



Laut Aufgabe 6, kann L daher auch nicht inhärent mehrdeutig sein. $\square$

Aufgabe 6. Zeigen Sie dass keine inhärent mehrdeutige reguläre Sprache existiert.

Hinweis: Transformieren Sie einen DFA in eine Grammatik.

Lösung. Sei $L \subseteq A^*$ eine reguläre Sprache und $D = \langle Q, A, \delta, q_0, F\rangle$ ein DFA für L, d.h. $\mathrm{L}(D) = L$. Definiere die Grammatik $G = \langle N, T, P, S\rangle$ durch $N = Q$, $T = A$,

$$P = \{q \to ap \mid \delta(q, a) = p\} \cup \{q \to \varepsilon \mid q \in F\},$$

und $S = q_0$. Dann behaupten wir, dass $\mathrm{L}(G) = \mathrm{L}(D) = L$. Das wurde im Beweis von Satz 2.1 bereits getan (für D als NFA). (Darüber hinaus, wurde gezeigt, dass G strikt rechtslinear ist.) Wir zeigen, dass G nun auch eindeutig ist, d.h. für alle $w \in L$ genau ein Ableitungsbaum existiert.
Sei $w = w_1 \cdots w_n$ mit $w_1, \dots, w_n \in A$ und

$$\delta(q_0, w_1) =: q_1,\ \delta(q_1, w_2) =: q_2,\ \dots,\ \delta(q_{n-1}, w_n) =: q_n.$$

Weil D ein DFA ist, ist δ eine Funktion, und $q_1, \dots, q_{n-1} \in Q$, $q_n \in F$ damit eindeutig. Laut Definition von P gilt nun

$$(q_0 \to w_1 q_1), (q_1 \to w_2 q_2), \dots, (q_{n-1} \to w_n q_n), (q_n \to \varepsilon) \in P.$$

Das können wir auch in Form einer Ableitung von w schreiben.

$$S \Longrightarrow_G w_1 q_1 \Longrightarrow_G w_1 w_2 q_2 \Longrightarrow_G \cdots \Longrightarrow_G w_1 \cdots w_n q_n \Longrightarrow_G w_1 \cdots w_n = w$$

Man sieht nun relativ leicht, dass diese Ableitung durch die Struktur von D und den Aufbau von w prädestiniert ist.

0. Sie muss jedenfalls, an der 0-ten Stelle, mit $q_0 = S$ anfangen.
1. Wenn $w = \varepsilon$, dann steht an der 1-ten Stelle ε und wir sind fertig. Weil G strikt rechtslinear ist, steht sonst an der 1-ten Stelle auf der linken Seite w_1. Laut der Definition von P, muss auf der rechten Seite $q_1 := \delta(q_0, w_1)$ stehen.
2. Wenn $w = w_1$, dann steht an der 2-ten Stelle w_1 und wir sind fertig. Weil G strikt rechtslinear ist, steht sonst an der 2-ten Stelle auf der linken Seite $w_1 w_2$. Laut der Definition von P, muss auf der rechten Seite $q_2 := \delta(q_1, w_2)$ stehen.

...

Dieser Algorithmus terminiert, weil $n < \infty$. Er liefert eine eindeutige Ableitung von w. $\square$