Prof. Javier Esparza
Philipp Czerner, Kevin Kappelmann, Valentin Krasotin

Technische Universität München
Lehrstuhl für Theoretische Informatik

Einführung in die Theoretische Informatik

Sommersemester 2024 – Hausaufgabenblatt 6

- Die Aufgaben werden in folgender Reihenfolge korrigiert: H6.3, H6.5, H6.2, H6.4.
- Die Knobelaufgabe bitte separat auf Moodle abgeben. Sie wird korrigiert.

AT-Aufgabe H6.1. (*Sabcxsaa*) unkorrigiert (6 Punkte)

Sei $\Sigma := \{a, b, x, y\}$, und sei $G = (V, \Sigma, P, S)$ eine kontextfreie Grammatik mit den Produktionen

$$S \to ABC \mid xS \mid AA \qquad A \to a \mid \varepsilon \mid C$$
$$B \to CBB \mid b \qquad C \to yC \mid x$$

Konvertieren Sie G mit dem aus der Vorlesung bekannten Verfahren schrittweise zur Chomsky-Normalform, indem Sie G so anpassen, dass für jede Produktion $\alpha \to \beta$ folgende zusätzlichen Einschränkungen gelten:

(a) $\beta \in \Sigma \cup V^*$ (b) $|\beta| \leq 2$ (c) $\beta \neq \varepsilon$ (d) $\beta \notin V$

Sie sollen also eine Grammatik G' angeben, die (a) und $L(G') = L(G)$ erfüllt, eine Grammatik G'', die (a), (b) und $L(G'') = L(G)$ erfüllt, usw. Geben Sie ihre Lösung für (d), also die Grammatik in CNF, auf Automata Tutor ab.

Hinweis: Nichtterminale können in AT nur aus einem einzigen Großbuchstaben bestehen. Nennen Sie also Nichtterminale wie X_a und X_{AB} passend um.

Lösungsskizze.

(a) Wir fügen $X \to x$ und $Y \to y$ hinzu, und ersetzten x und y in den anderen Produktionen jeweils durch X und Y.

$$S \to ABC \mid XS \mid AA \qquad A \to a \mid \varepsilon \mid C$$
$$B \to CBB \mid b \qquad C \to YC \mid x$$
$$X \to x \qquad Y \to y$$

(b) Wir passen jeweils die erste Produktion von S und von B an.

$$S \to AX_{BC} \mid XS \mid AA \qquad A \to a \mid \varepsilon \mid C$$
$$B \to CX_{BB} \mid b \qquad C \to YC \mid x$$
$$X \to x \qquad Y \to y$$
$$X_{BC} \to BC \qquad X_{BB} \to BB$$

(c) Zuerst suchen iterative die Nichtterminale, die ε erzeugen können: 1. $\{A\}$, 2. $\{A, S\}$ wegen $S \to AA$. Nun schauen wir uns jede Produktion an und erzeugen alle Produktionen, die durch das Entfernen von beliebigen A oder S Nichtterminalen entstehen können.

$$S \to AX_{BC} \mid XS \mid AA \mid X_{BC} \mid X \mid A \mid \varepsilon \qquad A \to a \mid \varepsilon \mid C$$
$$B \to CX_{BB} \mid b \qquad C \to YC \mid x$$
$$X \to x \qquad Y \to y$$
$$X_{BC} \to BC \qquad X_{BB} \to BB$$

Dann entfernen wir alle ε:

$$S \to AX_{BC} \mid XS \mid AA \mid X_{BC} \mid X \mid A \qquad A \to a \mid C$$
$$B \to CX_{BB} \mid b \qquad C \to YC \mid x$$
$$X \to x \qquad Y \to y$$
$$X_{BC} \to BC \qquad X_{BB} \to BB$$

(d) Wir haben im Moment die folgenden Kettenproduktionen:

$$S \to X_{BC} \mid X \mid A \qquad A \to C$$

Transitiv erhalten wir auch noch $S \to C$. Nach dem Entfernen ergibt sich also:

$$S \to AX_{BC} \mid XS \mid AA \mid BC \mid YC \mid x \mid a \qquad A \to a \mid YC \mid x$$
$$B \to CX_{BB} \mid b \qquad C \to YC \mid x$$
$$X \to x \qquad Y \to y$$
$$X_{BC} \to BC \qquad X_{BB} \to BB$$

Aufgabe H6.2. (*Der Wald vor lauter Bäumen*) 1+3+2 Punkte

Der kleine Theo sollte eigentlich seine Hausaufgaben machen, ist aber stattdessen in den Wald gegangen, um Bäume zu zählen. Können Sie ihm helfen?

Sei $G = (V, \{a\}, P, S)$ eine kontextfreie Grammatik mit genau $k \geq 1$ Nichtterminalen, sei $n \in \mathbb{N}$ mit $w = a^{n+1} \in L(G)$.

Hinweis: Die n-te Catalan-Zahl C_n ist (unter anderem) die Anzahl der vollen Binärbäume (d.h. alle Knoten haben entweder 2 oder 0 Kinder) mit $n+1$ Blättern. Sie dürfen die Catalan-Zahlen in Ihrer Lösung verwenden, ohne ihren Zahlenwert zu nennen.

(a) Nehmen wir an, dass die rechte Seite jeder Produktion höchstens 2 Zeichen lang ist. Wie viele Syntaxbäume hat w höchstens? Begründen Sie Ihre Antwort.

(b) Nehmen wir an, dass G in CNF ist. Wie viele Syntaxbäume hat w höchstens? Begründen Sie Ihre Antwort.

(c) Geben Sie eine Grammatik G' in CNF für die Sprache $\{a^3, a^8\}$ an. G' darf höchstens 5 Produktionen enthalten. Beachten Sie, dass z.B. $S \to AB \mid CD$ zwei Produktionen sind.

Lösungsskizze.

(a) Da man am Anfang eines Syntaxbaumes beliebig oft die Produktion $S \to S$ einfügen kann, gibt es für jedes Wort $w \in L(G)$ unendlich viele Syntaxbäume.

(b) Wenn man in einem Syntaxbaum von w alle Produktionen der Form $X \to a$ weglässt, erhält man einen vollen Binärbaum mit $n+1$ Blättern, in dem jeder Knoten mit einem Nichtterminal beschriftet ist. Da G in CNF ist, erhalten wir aus S ein Wort der Länge $n+1$ aus Nichtterminalen in genau n Schritten, und jeder Schritt fügt zwei neue Knoten zum Binärbaum hinzu. Insgesamt hat der Baum also $2n+1$ Knoten. Die Wurzel muss mit S beschriftet sein, für die anderen $2n$ Knoten gibt es jeweils höchstens k Möglichkeiten, jeder Binärbaum hat also höchstens k^{2n} Möglichkeiten für Beschriftungen. Laut Hinweis gibt es insgesamt C_n Binärbaume mit $n+1$ Blättern; insgesamt hat w also höchstens $k^{2n}C_n$ Syntaxbäume. Diese Obergrenze wird durch die Grammatik in CNF, die alle möglichen Produktionen enthält, erreicht.

(c)

$$\begin{aligned} S &\to A_4A_4 \mid A_2A_1 \\ A_4 &\to A_2A_2 \\ A_2 &\to A_1A_1 \\ A_1 &\to a \end{aligned}$$

Aufgabe H6.3. (*Matriririkelnununummern*) 5 Punkte

An der Technischen Hochschule Estlingen-Oberfeld werden Matrikelnummern nach einem kuriosen System vergeben: Matrikelnummern müssen Zahlen sein, deren Quersumme eine Primzahl ist, eine alte Estlinger Tradition. Nun hält der technische Wandel auch nicht vor Estlingen-Oberfeld ein, und der Rektor versucht verzweifelt, eine kontextfreie Grammatik zu finden, die alle korrekten Matrikelnummern erzeugt.

Sei $\Sigma := \{0, ..., 9\}$. Zeigen Sie, dass die Sprache $L := \{w \in \Sigma^* : \sum_{i\geq 1} w_i \text{ prim}\}$ nicht kontextfrei ist.

Lösungsskizze. Wir wenden das Pumping-Lemma für kontextfreie Sprachen an. Sei also $n \in \mathbb{N}$ beliebig. Wir wählen das Wort $z := 1^p$, wobei $p \geq n$ eine Primzahl ist. Nun erhalten wir eine Zerlegung $z = uvwxy$. Sowohl v als auch x enthalten nur das Zeichen 1, somit gilt $vx = 1^k$ für $k := |vx|$. Sei nun $i := 1 + p$. Wir erhalten

$$uv^iwx^iy = 1^{p+kp}$$

Da aber $p + kp = p(k+1)$ und $k > 0$ nach den Eigenschaften der Zerlegung gilt, besitzt $p(k+1)$ mindestens zwei Teiler (p und $k+1$), ist keine Primzahl, und $uv^iwx^iy \notin L$.

In beiden Fällen haben wir ein i gefunden, sodass $uv^iwx^iy \notin L$. Damit gilt die Eigenschaft des PL nicht, und L kann nicht kontextfrei sein.

Aufgabe H6.4. (*Die eine Deutung*) 7 Punkte

Zeigen Sie, dass die Grammatik $G := (\{E, T, F\}, \Sigma, P, E)$ mit $\Sigma = \{(,), 0, 1, +, *\}$ und P:

$$E \to T \mid E + T \qquad T \to F \mid T * F \qquad F \to 0 \mid 1 \mid (E)$$

(siehe Ü6.3) eindeutig, d.h. nicht mehrdeutig, ist.

Sie dürfen folgende Aussage ohne Beweis verwenden[1]:

$$\text{Jedes Wort in } L_G(E) \cup L_G(T) \cup L_G(F) \text{ ist wohlgeklammert.} \tag{1}$$

Lösungsskizze. Wir zeigen, dass G eindeutig ist, indem wir für jedes Nichtterminal $X \in \{E, T, F\}$ und jedes Wort $w \in L_G(X)$ einen eindeutigen ersten Schritt in einem Syntaxbaum von w, das mit X beginnt, angeben, d.h. eine eindeutige Produktion $X \to \beta$ und eindeutige Teilwörter von w für jedes Nichtterminal in β. Da ein Syntaxbaum eines Wortes $w \in L(G)$ mit E anfängt und jeder Schritt von einem Syntaxbaum zu Terminalen

[1] Ein Wort w ist wohlgeklammert, falls $|w|_(= |w|_)$ gilt und falls für alle Präfixe u von w $|u|_(\geq |u|_)$ gilt.

und zu Syntaxbäumen, die mit Nichtterminalen beginnen, führt, folgt daraus, dass der Syntaxbaum jedes Wortes eindeutig ist.

Sei $w \in L_G(F)$. Dann gilt entweder $w = 0$, $w = 1$ oder $w = (u)$ für ein $u \in \Sigma^*$. Der erste Schritt in einem Syntaxbaum von w, der mit F beginnt, ist also eindeutig.

Sei nun $w \in L_G(T)$. Wir unterscheiden zwei Fälle:

- Es gibt in w kein $*$, vor dem es genauso viele (wie) gibt. In diesem Fall kann die erste Produktion nicht $T \to T * F$ sein, denn dann gäbe es $u \in L_G(T)$, $v \in L_G(F)$ mit $w = u * v$, und laut (1) muss u genauso viele (wie) enthalten. Also muss die erste Produktion $T \to F$ sein. Das Teilwort, das von F erzeugt wird, ist eindeutig: es ist w.
- Es gibt in w ein $*$, vor dem es genauso viele (wie) gibt. Nehmen wir an, dass $w \in L_G(F)$, und seien $u, v \in \Sigma^*$ mit $w = u*v$. Dann gibt es $u', v' \in \Sigma^*$ mit $u = (u'$ und $v = v')$ und $u' * v' \in L_G(E)$. Da $u' * v'$ laut (1) wohlgeklammert ist, muss u' höchstens so viele) wie (enthalten, also muss u mehr (als) enthalten. Da das u mit $w = u * v$ beliebig war, folgt daraus, dass es kein $*$ in w gibt, vor dem es genauso viele (wie) gibt, ein Widerspruch. Also gilt $w \notin L_G(F)$, d.h. die erste Produktion muss $T \to T * F$ sein. Es bleibt noch zu zeigen, dass die Teilwörter von w, die von T und F erzeugt werden, eindeutig sind. Sei also $w = u * v$ mit $u \in L_G(T)$ und $v \in L_G(F)$. Wir wissen schon, dass u genauso viele (wie) enthält. Nehmen wir nun an, dass es $v_1, v_2 \in \Sigma^*$ gibt mit $v = v_1 * v_2$. Aus obigem Argument folgt, dass v_1 mehr (als) enthält, und somit enthält auch $u * v_1$ mehr (als). Da das $*$ in v beliebig war, folgt daraus, dass das $*$ in $u * v$ das letzte $*$ ist, vor dem es genauso viele (wie) gibt. Folglich sind die Teilwörter u und v eindeutig.

Für $w \in L_G(E)$ ist das Argument dasselbe wie für $w \in L_G(T)$, man ersetzt nur $*$ durch $+$ und $L_G(F)$ durch $L_G(T)$.

(Quiz-)Aufgabe H6.5. (*Auf den Hund gekommen*) 4+4 Punkte

Da Theo gerade zu beschäftigt damit ist, Bäume zu zählen, hat er seine Hausaufgaben von seinem Hund Trodo machen lassen. Leider hat Trodo die Aufgabenstellung nicht zu Ende gelesen, und ein paar wichtige Angaben nicht berücksichtigt. Theo hat aber eine Idee: vielleicht kann er die Grammatik von Trodo leicht anpassen und die Aufgabe so ohne viel Aufwand lösen.

Sei $\Sigma := \{a, b\}$ und $L \subseteq \Sigma^+$ eine kontextfreie Sprache. Konstruieren Sie eine kontextfreie Grammatik G' für die Sprachen

(a) L^a

(b) $\{w \in L : |w| \equiv 0 \pmod 7\}$

Hinweis: Sie können die Chomsky-Normalform verwenden und annehmen, dass eine Grammatik G in CNF mit $L(G) = L$ existiert.

Lösungsskizze. Sei $G = (V, \Sigma, P, S)$ in CNF mit $L(G) = L$. Wir konstruieren jeweils eine Grammatik $G' = (V', \Sigma, P', S')$.

(a) Als Variablen verwenden wir $V' := V \uplus \{X_a : X \in V\}$, und das Startsymbol ist nun $S' := S_a$. Die Idee ist, dass $L_{G'}(X_a) = L_G(X)^a$ für jedes $X \in V$ gilt, dass also X_a genau die Wörter erzeugt, die der Residualsprache der von X erzeugen Wörter entsprechen.

Die Produktionen P' konstruieren wir folgendermaßen:

- Alle Produktionen aus P sind in P' enthalten.

- Für $(X \to a) \in P$ fügen wir $X_a \to \varepsilon$ zu P' hinzu.
- Für $(X \to YZ) \in P$ fügen wir $X_a \to Y_a Z$ zu P' hinzu.

(b) Hier benutzen wir $V' := \{X_i : X \in V, i \in \{0, ..., 6\}\}$ und $S' := S_0$. Die Idee ist, dass $L_{G'}(X_i) = \{w \in L_G(X) : |w| \equiv i \pmod 7\}$ gilt, dass also X_i genau die Wörter erzeugt, die X erzeugt und deren Länge geteilt durch 7 Rest i hat. Dazu verwenden wir folgende Produktionen.

- Für $(X \to a) \in P$ fügen wir $X_1 \to a$ zu P' hinzu.
- Für $(X \to YZ) \in P$ fügen wir $X_i \to Y_j Z_k$ zu P' hinzu, für alle $i, j, k \in \{0, ..., 6\}$ mit $i \equiv j + k \pmod 7$.

Knobelaufgabe H6.6. (*Quadrat ist praktisch schlecht*)

Sei $n \in \mathbb{N}$, $\Sigma := \{1, ..., n\}$ und $L := \{w_1 w_2 ... w_k : w_1, w_2, ..., w_k \in \Sigma, w_1 < w_2 < ... < w_k\}$. Mit $n = 5$ gilt z.B. $\varepsilon, 12345, 145 \in L$ und $321, 141, 1335 \notin L$.

Zeigen Sie, dass es einen NFA N mit $L(N) = L$ gibt, sodass N genau $n + 1$ Zustände und höchstens $100n(\log_2 n)^2$ Transitionen hat. Sie dürfen hierfür einen NFA mit mehreren Startzuständen konstruieren.

Anmerkung: Diese Aufgabe ist schwierig. Falls keine korrekte Lösung abgegeben wird, bleibt die Musterlösung vielleicht geheim.