Formale Grundlagen der Informatik I 5. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Ulrich Kohlenbach
Alexander Kreuzer
Pavol Safarik

SS 2012

Gruppenübung

Aufgabe G1
Sei $\mathcal{A}$ ein DFA mit n Zuständen. Der Beweis des Pumping Lemmas zeigt dann, dass es für jedes Wort $x \in L(\mathcal{A})$ mit $|x| \geq n$ eine Zerlegung $x = u \cdot v \cdot w$ mit $|u \cdot v| \leq n$ und $|v| > 0$ gibt, so dass

$$u \cdot v^m \cdot w \in L(\mathcal{A})$$

für jedes $m \in \mathbb{N}$.

(a) Zeigen Sie, dass aus $L(\mathcal{A}) \neq \emptyset$ folgt, dass es ein Wort $x \in L(\mathcal{A})$ gibt mit $|x| < n$.
Hinweis: Betrachten Sie ein Wort $x \in L(\mathcal{A})$, das minimale Länge hat.

(b) Wie können Sie die Tatsache aus (a) benutzen um das Leerheitsproblem für reguläre Sprachen zu entscheiden?

(c) Zeigen Sie, dass $L(\mathcal{A})$ genau dann unendlich ist, wenn es ein Wort $x \in L(\mathcal{A})$ mit $n \leq |x| < 2n$ gibt.
Hinweis: Wenn die Sprache $L(\mathcal{A})$ unendlich ist, gibt es Wörter die zu $L(\mathcal{A})$ gehören und mindestens Länge $2n$ haben (warum?). Unter diesen, betrachten Sie ein Wort minimaler Länge.

Lösungsskizze:

(a) Sei x ein Wort in $L(\mathcal{A})$ von minimaler Länge. Wir zeigen, dass die Annahme $|x| \geq n$ zu einem Widerspruch führt: ist $|x| \geq n$, dann gibt es eine Zerlegung $x = u \cdot v \cdot w$, so dass $|u \cdot v| \leq n$, $|v| > 0$ und $u \cdot v^m \cdot w \in L(\mathcal{A})$ für jedes $m \in \mathbb{N}$. Insbesondere gehört für $m = 0$ auch $u \cdot w$ zur Sprache $L(\mathcal{A})$. Da aber $|v| > 0$, ist $|uw| < |x|$ und das widerspricht der Minimalität von x. Wir schließen, dass die Länge von x kleiner als n sein muss.

(b) Um zu entscheiden, ob die Sprache, die durch einen DFA mit n Zustände erkannt wird, leer ist, braucht man nur zu überprüfen, ob es ein Wort mit Länge kleiner als n gibt, das von diesem Automaten akzeptiert wird. Da es nur endlich viele solcher Wörter gibt, kann man einfach durch systematisches Durchsuchen feststellen, ob das zutrifft. Wenn es kein solches Wort gibt, ist die Sprache leer.

(c) $\Leftarrow$: Sei $x \in L(\mathcal{A})$ ein Wort mit $n \leq |x| < 2n$. Aufgrund des Pumping Lemmas gibt es eine Zerlegung $x = u \cdot v \cdot w$ mit $|v| > 0$ und $u \cdot v^m \cdot w \in L(\mathcal{A})$ für jedes $m \in \mathbb{N}$. Da $|v| > 0$, sind alle Wörter der Form $u \cdot v^m \cdot w$ verschieden, und deshalb ist die Sprache $L(\mathcal{A})$ unendlich.
$\Rightarrow$: Ist die Sprache $L(\mathcal{A})$ unendlich, dann gibt es Wörter in $L(\mathcal{A})$ von beliebiger Länge (einfach, weil die Anzahl von Wörtern bestimmter Länge endlich ist). Insbesondere gibt es Wörter, die mindestens

die Länge $2n$ haben. Sei $y \in L(\mathcal{A})$ von minimaler Länge unter der Nebenbedingung $|y| \geq 2n$ gewählt.

Wegen des Pumping Lemmas gibt es eine Zerlegung $y = u \cdot v \cdot w$, so dass $0 < |v| \leq n$ und $u \cdot w \in L(\mathcal{A})$. Sei $x = u \cdot w$. Da $x \in L(\mathcal{A})$ und $|x| = |y| - |v| < |y|$ gilt $|x| < 2n$ ($x \geq 2n$ würde der Minimalität von y widersprechen). Andererseits ist $|x| = |y| - |v| \geq 2n - n = n$. Deshalb ist x wie gewünscht.

Aufgabe G2

Sei

$$L = \{ss^{-1}t \,:\, s, t \in \{a,b\}^+\},$$

wobei s^{-1} die Umdrehung von s bezeichnet (wie auf dem letzten Übungsblatt definiert).

(a) Geben Sie eine kontextfreie Grammatik für L an.

(b) Zeigen Sie, dass L die Aussage im Pumping Lemma erfüllt.

(c) Zeigen Sie, dass L trotzdem nicht regulär ist.
Hinweis: Benutzen Sie den Satz von Myhill-Nerode!

Lösungsskizze:

(a) $G = (\{a,b\}, \{S, P, U\}, R, S)$

$$\begin{array}{rrcl} R: & S & \rightarrow & aPaUa \,|\, bPbUa \,|\, aPaUb \,|\, bPbUb \\ & P & \rightarrow & aPa \,|\, bPb \,|\, \varepsilon \\ & U & \rightarrow & Ua \,|\, Ub \,|\, \varepsilon \end{array}$$

(b) Wir zeigen, dass L die Behauptung des Pumping Lemmas für $n = 4$ erfüllt. Dafür sei $x = ss^{-1}t$ ein Wort, das zu L gehört, und mindestens Länge 4 hat. Es gibt zwei Möglichkeiten:

i. s hat Länge 1. Dann hat t mindestens Länge 2. Wir wählen $u := ss^{-1}$, also enthält u die ersten beiden Buchstaben von x. Wähle nun v als den dritten Buchstaben von x und w sei der Rest von x. Da u ein nicht-leeres Palindrom gerader Länge und w nicht leer ist, gilt $uv^mw \in L$ für alle $m \in \mathbb{N}$.

ii. s hat mindestens Länge 2. Dann wählen wir $u := \varepsilon$ und v der erste Buchstabe von x. w bezeichne wieder den (nicht leeren) Rest von x. Dann gehört $uv^mw = v^mw$ immer zu L: für $m > 1$ fängt v^mw nämlich mit vv an. Da v nur aus einem Buchstaben besteht, ist $vv = vv^{-1}$ und das Wort v^mw damit in L. Gilt $m = 0$, dann ist $v^mw = w$ und w gehört zu L, da $w = s_0s_0^{-1}vw$, wobei s_0 das Wort s ohne den ersten Buchstaben ist.

(c) Aus dem Satz von Myhill-Nerode folgt, dass es reicht, eine unendliche Menge von Wörtern zu finden, die bezüglich der Äquivalenzrelation $\sim_L$ paarweise inäquivalent sind. Betrachte die Wörter $\{ab^{2n+1}a \,:\, n \in \mathbb{N}\}$. Für $n \neq n'$ gilt $ab^{2n+1}a \not\sim_L ab^{2n'+1}a$, denn $ab^{2n+1}aw \in L$ und $ab^{2n'+1}aw \notin L$ für $w := ab^{2n+1}ab$.

Aufgabe G3

Betrachten Sie die kontextfreie Grammatik $G = (\{a,b\}, \{X_0, X, Y\}, P, X_0)$ mit

$$\begin{array}{rrcl} P: & X_0 & \rightarrow & aXaY \\ & X & \rightarrow & aXa \,|\, Y \\ & Y & \rightarrow & bY \,|\, \varepsilon. \end{array}$$

(a) Bestimmen Sie die von G erzeugte Sprache.

(b) Bestimmen Sie eine zu G äquivalente Grammatik G' ohne ε-Produktionen.

(c) Zeigen Sie, dass die Sprache

$$L = \{x \in \{a,b\}^* : |x|_a = |x|_b\}$$

kontextfrei ist, indem Sie eine kontextfreie Grammatik angeben, die diese Sprache erzeugt. (Begründen Sie Ihre Antwort!) Vgl. Übung 3.1.13 im Skript.

Lösungsskizze:

(a) Man zeigt mit Induktion, dass sich aus X genau a^nXa^n, $a^nb^mYa^n$ und $a^nb^ma^n$ erzeugen lassen. Daraus folgt $L(G) \subseteq \{a^nb^ma^nb^k : n \geq 1, m \geq 0, k \geq 0\}$. Umgekehrt kann man mit Induktion zeigen, dass aus Xdamit leicht eine Ableitung für jedes Wort aus dieser Menge angeben. Damit $L(G) = \{a^nb^ma^nb^k : n \geq 1, m \geq 0, k \geq 0\}$.

(b) Wir ersetzen $Y \to \varepsilon$ durch

$$\begin{array}{lcl} X_0 & \to & aXa \\ X & \to & \varepsilon \\ Y & \to & b. \end{array}$$

Und dann ersetzen wir $X \to \varepsilon$ durch

$$\begin{array}{lcl} X_0 & \to & aaY \\ X & \to & aa \\ X_0 & \to & aa. \end{array}$$

Insgesamt:

$$\begin{array}{lcl} X_0 & \to & aXaY \,|\, aXa \,|\, aaY \,|\, aa \\ X & \to & aXa \,|\, Y \,|\, aa \\ Y & \to & bY \,|\, b. \end{array}$$

(c)

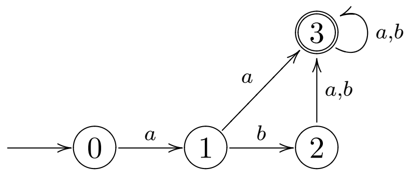
$$X \quad \to \quad \varepsilon \,|\, aXbX \,|\, bXaX$$

Es ist klar, dass jedes Wort w dieser Sprache $|w|_a = |w|_b$ erfüllt, da alle Produktionen diese Gleichheit erhalten (sie gilt also per Induktion für alle Zwischenergebnisse aller Ableitungen; insbesondere für die Endprodukte). Wir müssen nun begründen, warum ein beliebiges Wort w mit dieser Eigenschaft auch von der Grammatik erzeugt wird. Dazu benutzen wir Induktion über die Länge des Wortes w. Das leere Wort ε ist in L enthalten und kann mit unserer Grammatik erzeugt werden. Für den Induktionsschritt nehmen wir an, dass unsere Grammatik alle Wörter u mit $|u| < n$ und $|u|_a = |u|_b$ erzeugt. Sei nun w ein Wort der Länge n mit $|w|_a = |w|_b$. Wir nehmen an, dass das Wort mit a beginnt. Nun gibt es einen minimalen Präfix v von w mit $|v|_a = |v|_b$. Man macht sich leicht klar, dass dieser Präfix auf b enden muss, da es sonst einen kürzeren Präfix mit den gleichen Eigenschaften gäbe. Also gibt es eine Zerlegung $w = av'bu$, so dass v' und u in L enthalten sind (v' ist v ohne den ersten und letzten Buchstaben, demnach folgt $|v'|_a = |v'|_b$ und $|u|_a = |u|_b$ aus der Voraussetzung $|w|_a = |w|_b$). Nach Induktionsvorraussetzung können wir v' und u aus X mit unseren Ableitungsregeln erzeugen. Wenn wir also $X \to aXbX$ benutzen und dann aus dem ersten X v' und aus dem zweiten X u erzeugen, haben wir eine Ableitung für das Wort w konstruiert. Diese Argument gilt analog auch für Worte, die mit b beginnen, hier benutzt man die Produktion $X \to bXaX$.

Aufgabe G4

Sei $\Sigma = \{a, b\}$.

(a) Betrachten Sie den Automaten $\mathcal{A}$:



Geben Sie eine reguläre Grammatik an, die die Sprache $L(\mathcal{A})$ erzeugt.

(b) Sei L die Sprache, die von der regulären Grammatik $G = (\{a, b\}, \{S, T, U, V\}, P, S)$ erzeugt wird, wobei

$$\begin{array}{rlcl} P: & S & \to & aT \mid aU \mid bU \\ & T & \to & bT \mid bV \\ & U & \to & aV \mid \varepsilon \\ & V & \to & aV \mid bU. \end{array}$$

Geben Sie einen DFA an, der L erkennt.

(c) Zeigen sie dass die folgenden beiden Grammatiken G_1, G_2 mit Startsymbol X dieselbe reguläre Sprache erzeugen:

$$\begin{array}{rl} G_1: & X \to XaXaX \mid Y \\ & Y \to YY \mid b \mid \varepsilon \end{array} \qquad \begin{array}{rl} G_2: & X \to aY \mid bX \mid \varepsilon \\ & Y \to bY \mid aZ_1 \mid aX \\ & Z_1 \to aZ_2 \mid bX \mid \varepsilon \\ & Z_2 \to bY \mid aZ_1 \mid aX \end{array}$$

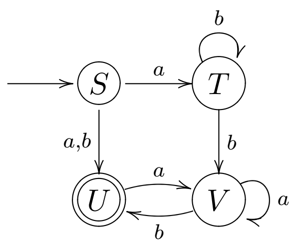
Dazu übersetzt man zunächst die zweite, rechtslineare Grammatik in einen (möglichst kompakten) DFA um die erzeugte Sprache zu identifizieren. Dann zeige man, dass die erste Grammatik ebenfalls genau diese Sprache erzeugt (mit Induktionsbeweisen für die beiden Inklusionen!).

Lösungsskizze:

(a) $G = (\{a, b\}, \{X_0, X_1, X_2, X_3\}, P, X_0)$

$$\begin{array}{rlcl} P: & X_0 & \to & aX_1 \\ & X_1 & \to & aX_3 \mid bX_2 \\ & X_2 & \to & aX_3 \mid bX_3 \\ & X_3 & \to & aX_3 \mid bX_3 \mid \varepsilon \end{array}$$

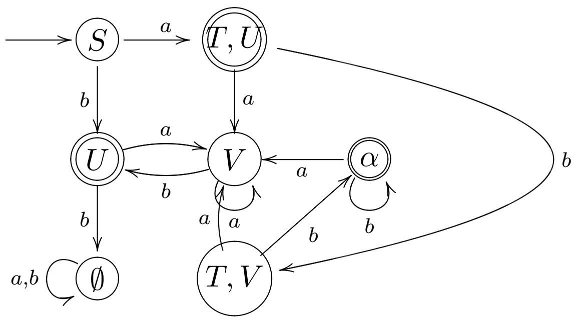
(b) Der folgende NFA erkennt L:



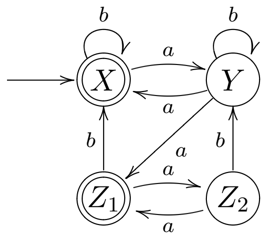
Wir determinisieren diesen Automaten um einen DFA zu bekommen, der L erkennt:

δ	a	b
$\{S\}$	$\{T, U\}$	$\{U\}$
$\{T, U\}$	$\{V\}$	$\{T, V\}$
$\{U\}$	$\{V\}$	$\emptyset$
$\{V\}$	$\{V\}$	$\{U\}$
$\{T, V\}$	$\{V\}$	$\{T, V, U\}$
$\{T, V, U\}$	$\{V\}$	$\{T, V, U\}$
$\emptyset$	$\emptyset$	$\emptyset$

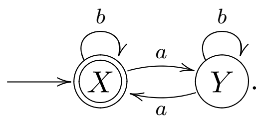
Akzeptierend Zustände sind $\alpha := \{T, V, U\}, \{T, U\}$ und $\{U\}$. In einem Bild:



(c) Da G_2 rechtslinear ist, lässt sie sich direkt in folgenden NFA übersetzten:



Nach einer Determinisierung und Minimierung erhält man folgenden DFA:



Alternativ kann die Grammatik G_2 auch vorher vereinfacht werden. Dafür kennen wir aber *kein* Verfahren, das in jedem Fall funktioniert: Z_2 und Y können miteinander identifiziert werden, da sie das Gleiche produzieren. Nach dieser Identifikation sieht man, dass auch Z_1 und X identifiziert werden können. Wir erhalten folgende Grammatik

$$X \to aY \mid bX \mid \varepsilon$$
$$Y \to bY \mid aX$$

und auch den obigen DFA.

Die Grammatiken erzeugen damit die Sprache $L := \{x \in \{a, b\}^* \colon |x|_a \equiv_2 0\}$.

Mit Induktion sieht man leicht, dass die Sprache der ersten Grammatik in L enthalten ist.

Wir zeigen mit Induktion über $|w|_a/2$, dass mit der ersten Grammatik alle Worte aus L erzeugt werden können.

Induktionsanfang: Falls $|w|_a/2 = 0$, dann besteht w nur aus bs und kann offensichtlich erzeugt werden.

Induktionsschritt: Angenommen alle Worte $v \in L$ mit $|v|_a/2 = n$ können erzeugt werden. Sei $w \in L$ nun ein Wort mit $|w|_a/2 = n + 1$. Dann lässt sich w schreiben als

$$w = uau'av$$

mit $u, u' \in L(b^*)$ und $v \in L$ und $|v|_a/2 = n$. u, u' können trivial aus X produziert werden und v nach Induktionsvoraussetzung. Damit produziert $XaXaX$ das Wort w.

Hausübung

Aufgabe H1 (6 Punkte)

Betrachten Sie die folgende kontextfreie Grammatik $G_a = (\Sigma, V_a, P_a, S)$ über dem Alphabet $\Sigma = \{a, b\}$, mit der Variablenmenge $V_a = \{S, T\}$, dem Startsymbol S und der Produktionen P_a gegeben durch

$$\begin{aligned} S &\rightarrow\ aSbT \mid ab \\ T &\rightarrow\ bTa \mid \varepsilon \end{aligned}$$

Formen Sie die Grammatik G_a in eine Grammatik G in Chomsky-Normalform um, so dass $L(G_a) = L(G)$. Geben Sie wesentliche Zwischenschritte an.