Fachbereich Mathematik
Achim Blumensath
Benno van den Berg

TECHNISCHE UNIVERSITÄT DARMSTADT

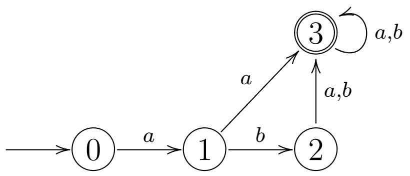
20. Mai, 2009

5. Übungsblatt Formale Grundlagen der Informatik I
SS 2009

(E5.1)

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

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



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

(b) Welche Sprache wird von der folgenden Grammatik erzeugt?

$$\begin{array}{lcl} X_0 & \to & abX_0 \mid \varepsilon \\ ab & \to & ba \\ ba & \to & ab \end{array}$$

(c) Nennen wir die Sprache die von der Grammatik in (b) erzeugt wird, L. Zeigen Sie, dass die Sprache L kontextfrei ist, indem Sie eine kontextfreie Grammatik angeben, die diese Sprache erzeugt. (Und begründen Sie Ihre Antwort!)

Musterlösung.

(a)

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

(b) Die erzeugte Sprache ist $\{x \in \Sigma^* : |x|_a = |x|_b\}$.

(c)

$$X_0 \rightarrow \varepsilon \mid X_0X_0 \mid aX_0b \mid bX_0a$$

(E5.2) [Chomsky-Normalform]

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

$$\begin{array}{rlcl} P: & X_0 & \rightarrow & aXY \mid bYb \mid a \\ & X & \rightarrow & aXa \mid b \\ & Y & \rightarrow & bX_0a \mid aX_0 \mid X \end{array}$$

Konstruieren Sie eine zu G äquivalente Grammatik in Chomsky-Normalform.

Musterlösung.

1. Schritt (Variablen vor Buchstaben):

$$\begin{array}{rlcl} P: & X_0 & \rightarrow & Z_aXY \mid Z_bYZ_b \mid Z_a \\ & X & \rightarrow & Z_aXZ_a \mid Z_b \\ & Y & \rightarrow & Z_bX_0Z_a \mid Z_aX_0 \mid X \\ & Z_a & \rightarrow & a \\ & Z_b & \rightarrow & b \end{array}$$

2. Schritt (elimiere $X \rightarrow Y$):

$$\begin{array}{rlcl} P: & X_0 & \rightarrow & Z_aXY \mid Z_bYZ_b \mid a \\ & X & \rightarrow & Z_aXZ_a \mid b \\ & Y & \rightarrow & Z_bX_0Z_a \mid Z_aX_0 \mid Z_aXZ_a \mid b \\ & Z_a & \rightarrow & a \\ & Z_b & \rightarrow & b \end{array}$$

3. Schritt (eliminiere $X \rightarrow X_0 \ldots X_k$ mit $k \geqslant 3$):

$$\begin{array}{rlcl} P: & X_0 & \rightarrow & SY \mid TZ_b \mid a \\ & X & \rightarrow & SZ_a \mid b \\ & Y & \rightarrow & UZ_a \mid Z_aX_0 \mid SZ_a \mid b \\ & S & \rightarrow & Z_aX \\ & T & \rightarrow & Z_bY \\ & U & \rightarrow & Z_bX_0 \\ & Z_a & \rightarrow & a \\ & Z_b & \rightarrow & b \end{array}$$

(E5.3)

Welche der folgenden Sprachen sind (i) regulär, (ii) kontexfrei, aber nicht regulär, oder (iii) nicht kontextfrei? Begründen Sie Ihre Antwort!

$$\begin{aligned} L_1 &= \{x \in \{a,b\}^* : |x|_a \geqslant |x|_b\} \\ L_2 &= \{x \in \{a,b,c\}^* : |x|_a \geqslant |x|_b \geqslant |x|_c\} \\ L_3 &= \{x \in \{a,b\}^* : |x|_a \geqslant |x|_b \text{ und } |x|_b \leqslant 2009\} \\ L_4 &= \{x \in \{a,b\}^* : |x|_a \geqslant |x|_b \text{ und } |x|_b \geqslant 2009\} \end{aligned}$$

Geben Sie jeweils eine Grammatik an, die die Sprache erzeugt.

Musterlösung.

L_1: L_1 ist kontextfrei, aber nicht regulär. Eine kontextfreie Grammatik für L_1 (mit Startsymbol S) wäre

$$P: \quad X_0 \quad \rightarrow \quad \varepsilon \mid X_0X_0 \mid aX_0b \mid bX_0a \mid aX_0 \mid X_0a$$

L_1 ist nicht regulär. Angenommen L_1 wäre regulär. Sei n die Konstante aus dem Pumping Lemma und setze $x = a^n b^n$. Nach dem Lemma gibt es eine Zerlegung $x = uvw$ mit $|uv| \leqslant n, v \neq \varepsilon$ und $uv^m w \in L_1$ für alle $m \in \mathbb{N}$. Da $|uv| \leqslant n$ und $v \neq \varepsilon$, ist $v = a^k$ für $k \geqslant 1$. Also ist $uv^0w = a^{n-k}b^n \notin L_1$. Widerspruch!

L_2: L_2 ist nicht kontextfrei. Angenommen L_2 wäre kontextfrei. Sei n die Konstante aus dem Pumping Lemma und setze $x = a^n b^n c^n$. Nach dem Lemma gibt es eine Zerlegung $x = yuvwz$ mit $|uvw| \leqslant n, uw \neq \varepsilon$ und $yu^m vw^m z \in L_2$ für alle $m \in \mathbb{N}$. Da $|uvw| \leqslant n$ kann uvw nicht sowohl a's als auch c's enthalten. Deshalb gibt es nur die folgenden Möglichkeiten:

- u und w enthalten nur a. Dann enthält yu^0vw^0z weniger a's als b's und ist deshalb nicht in L_2 enthalten. Widerspruch!
- u und w enthalten a und b. Dann enthält yu^0vw^0z weniger b's als c's und ist deshalb nicht in L_2 enthalten. Widerspruch!
- u und w enthalten nur b. Dann enthält yu^0vw^0z weniger b's als c's und ist deshalb nicht in L_2 enthalten. Widerspruch!
- u und w enthalten b und c. Dann enthält yu^2vw^2z mehr b's als a's und ist deshalb nicht in L_2 enthalten. Widerspruch!
- u und w enthalten nur c. Dann enthalt yu^2vw^2z mehr c's als a's und ist deshalb nicht in L_2 enthalten. Widerspruch!

Eine Grammatik für L_2 ist:

$$\begin{aligned} X_0 &\rightarrow abX_0 \mid abcX_0 \mid aX_0 \mid \varepsilon \\ ab &\rightarrow ba \\ ba &\rightarrow ab \\ bc &\rightarrow cb \\ cb &\rightarrow bc \\ ac &\rightarrow ca \\ ca &\rightarrow ac \end{aligned}$$

L_3: L_3 ist regulär. Für jedes $n \in \mathbb{N}$ ist

$$L_3^n = \{x \in \Sigma^* \,:\, |x|_a \geqslant n \text{ und } |x|_b = n\} = \{x \in \Sigma^* \,:\, |x|_a \geqslant n\} \cap \{x \in \Sigma^* \,:\, |x|_b = n\}$$

ein Durchschnitt von zwei regulären Sprachen und deshalb regulär. Es folgt, dass $L_3 = \bigcup_{n \leqslant 2009} L_3^n$ auch regulär ist.

Man kann auch zeigen, dass L_3 regulär ist, indem man einen DFA angibt, der L_3 erkennt. Sei $\mathcal{A} = (\Sigma, Q, q_0, \delta, A)$ wobei $\Sigma = \{a, b\}$, $Q = \{q_{i,j} \,:\, 0 \leqslant i \leqslant 2010, 0 \leqslant j \leqslant 2010\}$, $q_{0,0}$ der Anfangszustand ist,

$$\begin{aligned} \delta(q_{i,j}, a) &= \begin{cases} q_{i+1,j} & \text{falls } i < 2010 \\ q_{i,j} & \text{falls } i = 2010 \end{cases} \\ \delta(q_{i,j}, b) &= \begin{cases} q_{i,j+1} & \text{falls } j < 2010 \\ q_{i,j} & \text{falls } j = 2010 \end{cases} \end{aligned}$$

(also die erste Komponente i zählt die as, und die zweite Komponente j die bs) und $A = \{q_{i,j} \,:\, i \geqslant j, j \neq 2010\}$. Dieser DFA erkennt L_3 und kann, wie in E5.1(a), in einer regulären Grammatik umgewandelt werden.

L_4: L_4 ist kontextfrei, aber nicht regulär. Wäre L_4 regulär, dann wäre $L_1 = L_3 \cup L_4$ das auch. Wir haben aber oben gezeigt, dass L_1 nicht regulär ist.

Eine kontextfreie Grammatik für L_4 wäre, mit Variablen $\{X_i \,:\, 0 \leqslant i \leqslant 2009\} \cup \{S_i \,:\, 0 \leqslant i \leqslant 2009\}$, Startsymbol S_{2009}:

$$\begin{aligned} X_i &\rightarrow \{X_p X_q \,:\, p + q = i\} \mid aX_{i-1}b \mid bX_{i-1}a \mid aX_0 \mid X_0 a \quad \text{für } 0 < i \leqslant 2009 \\ X_0 &\rightarrow \varepsilon \mid X_0 X_0 \mid aX_0 b \mid bX_0 a \mid aX_0 \mid X_0 a \end{aligned}$$

(Die Idee ist, dass X_i durch ein Wort ersetzt wird, das mindestens i bs enthält.)

Hausaufgabe

(H5.4)

Sei $L = \{a^n b^m a^n c^{n+m} \,:\, m, n \in \mathbb{N}\}$.

(a) Zeigen Sie, dass L nicht kontextfrei ist.

(b) Geben Sie eine Grammatik an, die die Sprache L erzeugt.

Musterlösung.

(a) Wir zeigen, dass das Pumping Lemma verletzt ist. Das heißt:

Für jedes $i \in \mathbb{N}$ gibt es ein $x \in L$ mit $|x| \geqslant i$, so dass für alle Zeichenreihen y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leqslant i$, es ein $j \in \mathbb{N}$ gibt mit

$$y \cdot u^j \cdot v \cdot w^j \cdot z \notin L.$$

Sei $i \in \mathbb{N}$ beliebig und betrachte

$$x = a^i b^i a^i c^{2i}.$$

Offensichtlich ist $x \in L$ und $|x| \geqslant i$. Seien also y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leqslant i$. Wir wählen $j = 2$ und behaupten

$$x' = y \cdot u^2 \cdot v \cdot w^2 \cdot z \notin L.$$

Beweis: Nenne die Teilwörter der Form a^i, b^i und c^{2i} in x *Blöcke*. Es gibt jetzt zwei Möglichkeiten:

- u oder w enthält as. Wegen $|uvw| \leqslant i$ müssen die as die in u und w enthalten sein zum selben Block gehören. Das bedeutet, dass x' nicht zwei gleich grosse a-Blöcke hat und nicht zu L gehören kann.
- u und w enthalten keine as. Dann sind u und w entweder beide in dem b-Block enthalten oder beide in dem c-Block. Wenn beide in dem b-Block enthalten sind, dann hat x' mehr als i bs, weil die beide a-Blöcke Grosse i haben, und das c-Block Grosse i. Deshalb $x' \notin L$. Andererseits, wenn beide in dem c-Block enthalten sind, dann hat x' mehr als $2i$ cs, weil die beide a-Blöcke Grosse i haben, und das b-Block Grosse i. Also auch dann $x' \notin L$.

(b)

$$\begin{array}{lcl}
X_0 & \to & PaXacQ \mid \varepsilon \mid bYc \\
X & \to & aXac \mid \varepsilon \mid bYc \\
Y & \to & bYc \mid \varepsilon \\
ca & \to & ac \\
Pa & \to & aP \\
Pb & \to & bP \\
Pc & \to & cR \\
Rc & \to & cR \\
RQ & \to & \varepsilon
\end{array}$$