Fachbereich Mathematik
Prof. Dr. Martin Otto
Alexander Kartzow
Alexander Kreuzer
Benno van den Berg

TECHNISCHE UNIVERSITÄT DARMSTADT

SS 2010
26.05.2010

7. Übungsblatt zu FGdI 1

Gruppenübung

Aufgabe G1

Zeigen Sie, dass die Klasse der rekursiv aufzählbaren Σ-Sprachen unter Durchschnitt und Vereinigung abgeschlossen ist.
Folgerns sie daraus, dass auch die Klasse der entscheidbaren Σ-Sprachen unter Durchschnitt und Vereinigung abgeschlossen sind.
Hinweis: Sie können dafür die Tatsache benutzten, dass es für zwei beliebige Turingmaschinen immer eine Turingmaschine gibt, die diese beiden parallel simuliert. Diese Turingmaschine kann man aus der Beschreibung der beiden anderen explizit konstruieren (was aufwendig ist). Unter Annahme der Church-Turing-These ist klar, dass eine solche Konstruktion existieren muss, denn auch ein Computer kann die Ausführung von zwei Programmen parallel simulieren.

Musterlösung:
Nach Definition 4.3.2 heißt $L \subseteq \Sigma^*$ aufzählbar, wenn L von einer Turingmaschine akzeptiert wird. Nehmen wir deshalb an, dass $K \subseteq \Sigma^*$ von einer Turingmaschine $\mathcal{M}_K$ und $L \subseteq \Sigma^*$ von einer Turingmaschine $\mathcal{M}_L$ akzeptiert werden.
Die folgende Maschine $\mathcal{M}$ akzeptiert die Sprache $K \cap L$: auf einer beliebigen Eingabe $w \in \Sigma^*$ simuliert die Maschine $\mathcal{M}$ gleichzeitig die Berechnungen von $\mathcal{M}_K$ und $\mathcal{M}_L$ auf dieser Eingabe. Wenn beide Simulationen in akzeptierenden Zuständen terminieren, dann akzeptiert $\mathcal{M}$ die Eingabe. Andernfalls gilt die Eingabe als verworfen. (Dies gilt insbesondere wenn eine der beiden Simulationen nicht terminiert).
Die Sprache $K \cup L$ wird von der folgenden Maschine $\mathcal{M}'$ akzeptiert: Auch diese Maschine simuliert die beiden Maschinen $\mathcal{M}_K$ und $\mathcal{M}_L$, nur die Akzeptanzbedingung ändert sich. Sobald eine der beiden Simulationen in einem akzeptierenden Zustand terminiert, terminiert auch $\mathcal{M}'$ und akzeptiert die Eingabe. Wenn dieser Fall nie eintritt, gilt die Eingabe als verworfen.
Um die entsprechen Abschlüsse für entscheidbare Sprachen zu folgern wenden wir Beobachtung 4.3.3. aus dem Skript an: Sind $K, L \subseteq \Sigma^*$ entscheidbare Sprachen, dann ist $K, \overline{K}, L, \overline{L}$ aufzählbar. Nach dem oben gezeigten gilt dann auch, dass $K \cap L$ und $\overline{K \cap L} = \overline{K} \cup \overline{L}$ aufzählbar sind und damit dass $K \cap L$ entscheidbar ist.
Die Entscheidbarkeit von $K \cup L$ kann man analog zeigen.

Aufgabe G2

Bestimmen Sie eine kontextfreie Grammatik für die arithmetischen Terme über $(\mathbb{N}, +, \cdot, 0, 1)$, die minimal geklammert sind (d.h., wir berücksichtigen die Assoziativität der Addition und Multiplikation und geben der Multiplikation Priorität über die Addition).

Musterlösung: Wir setzen $\Sigma = \{0, 1, +, \cdot, (,)\}$ und benutzen die Nichtterminale $\{S, B, A, M\}$. S ist das Startsymbol, B steht für Terme die nur aus einer der Konstanten 0 oder 1 bestehen, A steht für alle Terme, die Summen kleinerer Terme sind, und M steht für alle Terme, die

Produkte kleinerer Terme sind. Wir benutzen die folgenden Produktionen:

$$
\begin{array}{lcl}
S & \to & B \mid A \mid M \\
B & \to & 0 \mid 1 \\
A & \to & A + A \mid A + B \mid B + A \mid B + B \mid M + A \mid A + M \mid M + M \mid M + B \mid B + M \\
M & \to & M \cdot M \mid M \cdot B \mid B \cdot M \mid B \cdot B \mid M \cdot (A) \mid (A) \cdot M \mid (A) \cdot (A) \mid (A) \cdot B \mid B \cdot (A)
\end{array}
$$

Aufgabe G3

Sei w ein Wort über einem Alphabet Σ und

$$L_w := \Sigma^* \cdot w$$

die Sprache der Wörter, die auf w enden. Skizzieren Sie ein Verfahren, dass zu jedem Wort $w \in \Sigma^*$ den minimalen DFA $\mathcal{A}_w$ für L_w liefert.

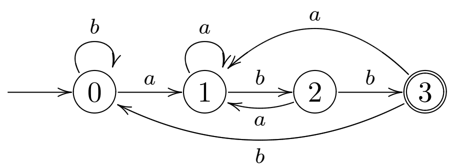
Bemerkung: Die effiziente Berechnung und Simulation des Minimalautomaten für L_w ist die Grundlage für den sehr effizienten String-Matching-Algorithmus von Knuth, Morris und Pratt.

Hinweis 1: Versuchen Sie erst den minimalen DFA für L_w zu bestimmen im Fall $\Sigma = \{a, b\}$ und $w = abb$ und $w = ababb$.

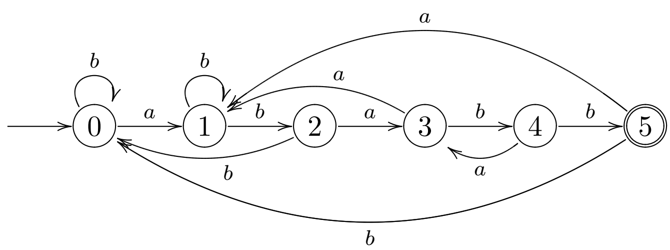
Hinweis 2: Lesen Sie Paragraph 5.1 im Skript (Seiten 72 und 73).

Musterlösung:

Der minimale Automat für L_{abb} ist:



Der minimale Automat für L_{ababb} ist:



Im allgemeinen ist der minimale Automat für L_w der DFA $\mathcal{A}_w = (\Sigma, Q, q_0, \delta, A)$, wobei

$$
\begin{array}{rcl}
Q & = & \{w' \,:\, w' \text{ ist ein Prefix von } w\} \\
q_0 & = & \epsilon \\
\delta(w', x) & = & w'', \text{ falls } w'' \text{ das längste Prefix von } w \text{ ist,} \\
 & & \text{das gleichzeitig ein Suffix von } w'x \text{ ist} \\
A & = & \{w\}
\end{array}
$$

Beweisskizze: Wir zeigen erst, dass der DFA $\mathcal{A}_w$ die Sprache L_w erkennt. Dazu beweist man mit Induktion über die Länge des Wortes, dass für alle $v \in \Sigma^*$ gilt, dass $\hat{\delta}(q_0, v)$ das längste Prefix von w ist, das gleichzeitig ein Suffix von v ist. Deshalb gilt:

$$\hat{\delta}(q_0, v) \in A \quad \text{gdw.} \quad \hat{\delta}(q_0, v) = w \quad \text{gdw.} \quad w \text{ ist ein Suffix von } v.$$

Um die Minimalität von $\mathcal{A}_w$ zu zeigen, benutzen wir das $L_w \neq \emptyset$. Wenn es einen Automaten mit n Zuständen gibt, der L_w erkennt, dann gibt es ein Wort $v \in L_w$ mit $|v| < n$ (vgl. Hausaufgabe H2 auf Blatt 5). Da das kürzeste Wort in L_w das Wort w ist, muss also jeder DFA für L_w mindestens $|w| + 1$ viele Zustände haben. Da $|Q| = |w| + 1$, ist $\mathcal{A}_w$ also minimal.

Aufgabe G4

Sei $L \subseteq \{a\}^*$ eine kontextfreie Sprache über einem einelementigen Alphabet.

(a) Zeigen Sie, dass man L darstellen kann als endliche Vereinigung

$$L = L_0 \cup L_{k_1,p_1} \cup \dots \cup L_{k_m,p_m} ,$$

wobei L_0 eine endliche Sprache ist und

$$L_{k,p} := \{ \, a^{k+ip} : i \in \mathbb{N} \, \} .$$

(b) Folgern Sie hieraus, dass jede kontextfreie Sprache über einem einelementigen Alphabet regulär ist.

Hinweis zu (a): Benutzen Sie das Pumping Lemma um zu zeigen, dass es eine Zahl $n \in \mathbb{N}$ gibt, so dass

$$L = L_0 \cup L_{k_1,p_1} \cup L_{k_2,p_2} \cup \dots \quad \text{mit } p_1, p_2, \dots \leq n ,$$

wobei Sie zunächst unendlich viele Sprachen der Form $L_{k,p}$ zulassen. Argumentieren Sie im zweiten Schritt, wieso Sie mit endlich vielen solchen Sprachen auskommen.

Musterlösung:

(a) Sei n die Konstante aus dem Pumping Lemma. Wir setzen

$$\begin{aligned} L_0 &:= \{ \, w \in L : |w| < n \, \} \text{ und} \\ S &:= \{ k \in \mathbb{N} \, : \, a^k \in L \text{ und } k \geq n \}. \end{aligned}$$

Für jedes $k \in S$ gibt es eine Zerlegung $a^k = yuvwz$ mit $|uvw| \leq n$, $uw \neq \varepsilon$ und $yu^ivw^iz \in L$ für alle $i \in \mathbb{N}$. Setze $p_k := |uw|$. Dann folgt, dass

$$yu^ivw^iz = a^{k+ip_k} \in L .$$

Also ist $L_{k,p_k} \subseteq L$ und

$$L = L_0 \cup \bigcup_{k \in S} L_{k,p_k} .$$

Sind $k, l \in S$ zwei Zahlen mit

$$p_k = p_l \, , \quad k \equiv_{p_k} l \quad \text{und} \quad k < l \, ,$$

dann gilt $L_{l,p_l} \subseteq L_{k,p_k}$. Wir können also L_{l,p_l} aus der Vereinigung entfernen. Deswegen gilt

$$L = L_0 \cup \bigcup_{k \in S_0} L_{k,p_k} ,$$

wobei

$$S_0 := \left\{ k \in S \, : \, \text{es gibt kein } l \in S \text{ mit } l < k \, , \; p_l = p_k \text{ und } k \equiv_{p_k} l \right\} .$$

Da $p_k \leq n$ ist, folgt, dass S_0 weniger als n^2 Elemente enthält. Insbesondere ist die obige Vereinigung also endlich.

(b) Endliche Sprachen und jede Sprache der Form $L_{k,p}$ sind regulär:

$$L_{k,p} = L(a^k(a^p)^*) .$$

Da reguläre Sprachen unter Vereinigung abgeschlossen sind, folgt, dass

$$L = L_0 \cup L_{k_1,p_1} \cup \dots \cup L_{k_n,p_n}$$

ebenfalls regulär ist.

Hausübung

Aufgabe H1 (7 Punkte)

Wahr oder falsch? Begründen Sie Ihre Antwort!

(a) Reguläre Sprachen sind entscheidbar.

(b) Es gibt kontextfreie Sprachen, die regulär sind.

(c) Ist L_1 regulär und $L_2 \subseteq L_1$, so ist L_2 auch regulär.

(d) Ist L_1 regulär und L_2 beliebig, dann ist

$$L = \{x \in \Sigma^* : \text{es existiert ein } y \in L_2, \text{ so dass } xy \in L_1\}$$

regulär.

(e) Das Komplement einer Sprache, die von einer Grammatik erzeugt wird, wird auch wieder von einer Grammatik erzeugt.

(f) Sind L_1 und L_2 kontextfrei, so ist $L_1 \setminus L_2$ entscheidbar.

(g) Die Sprache aller regulären Ausdrücke ist regulär.

Musterlösung:

(a) Richtig. (Jede kontextsensitive Sprache ist entscheidbar und jede reguäre Sprache ist auch eine kontextsensitive Sprache.)

(b) Richtig. (Jede reguläre Sprache ist auch kontextfrei.)

(c) Falsch. (Sei $L_1 = \Sigma^*$ und L_2 eine nicht-reguläre Σ-Sprache.)

(d) Richtig. (Sei $\mathcal{A}$ ein DFA für die Sprache L_1. Wir können $\mathcal{A}$ in einem DFA $\mathcal{A}'$ für die Sprache L umwandeln, indem wir die Menge der akzeptierenden Zustände ändern: wir erklären nämlich diejenigen Zustände in $\mathcal{A}'$ zu akzeptierenden, aus denen man mit einem Wort $y \in L_2$ in einen akzeptierenden Zustand von $\mathcal{A}$ gelangt. Diese Definition ist natürlich nicht unbedingt rekursiv, d.h., es gibt Sprachen L_2, so dass der Automat für L sich nicht aus dem Automaten für L_1 berechnen lässt).

(e) Falsch. (Die Sprachen, die von Grammatiken erzeugt werden, sind gerade die rekursiv aufzählbaren Sprachen. Diese sind nicht unter Komplement abgeschlossen. Rekursiv aufzählbare Mengen, deren Komplement auch rekursiv aufzählbar ist, sind gerade die entscheidbaren Mengen: siehe Beobachtung 4.3.3 auf Seite 66 im Skript.)

(f) Richtig. (Kontextfrei Sprachen sind entscheidbar und entscheidbare Sprachen sind unter allen Booleschen Operationen abgeschlossen.)

(g) Falsch. (Wäre die Sprache aller regulären Ausdrücken regulär, dann wäre die Sprache der Klammerwörter auch regulär (warum?). Das ist sie aber nicht.)

Aufgabe H2 (8 Punkte)

Welche der folgenden Sprachen über dem Alphabet $\Sigma = \{a, b, c\}$ sind (i) regulär, (ii) kontexfrei, aber nicht regulär, oder (iii) nicht kontextfrei? Begründen Sie Ihre Antwort!

$$\begin{aligned}
L_1 &= \{x \in \Sigma^* : |x|_a > |x|_b\} \\
L_2 &= \{x \in \Sigma^* : |x|_a > |x|_b > |x|_c\} \\
L_3 &= \{x \in \Sigma^* : |x|_a > |x|_b \text{ und } |x|_b \leq 2010\} \\
L_4 &= \{x \in \Sigma^* : |x|_a > |x|_b \text{ und } |x|_b \geq 2010\}
\end{aligned}$$

Musterlösung:

L_1: L_1 ist kontextfrei, aber nicht regulär. Ein PDA für L_1 ist $\mathcal{P} = (\Sigma, Q, q_0, \Delta, A, \Gamma, \#)$ mit $Q = \{q_0, q_1\}$, $\Gamma = \{\#, A, B\}$, $A = \{q_1\}$ und Transitionen

$$\begin{array}{llllll} \{ & (q_0, & \#, & a, & A\#, & q_0) \\ & (q_0, & \#, & b, & B\#, & q_0) \\ & (q_0, & A, & a, & AA, & q_0) \\ & (q_0, & B, & a, & \epsilon, & q_0) \\ & (q_0, & A, & b, & \epsilon, & q_0) \\ & (q_0, & B, & b, & BB, & q_0) \\ & (q_0, & A, & \epsilon, & \epsilon, & q_1) \\ & (q_1, & A, & \epsilon, & \epsilon, & q_1) \\ & (q_1, & \#, & \epsilon, & \epsilon, & q_1) \quad \}. \end{array}$$

Man kann auf (mindestens) zwei Weisen argumentieren, warum L_1 nicht regulär ist.

Wenn L_1 regulär ist, dann auch die Sprachen $L_5 := \{w \in \Sigma^* : |w|_a \leq |w|_b\}$ (da reguläre Sprachen unter Komplementbildung abgeschlossen sind) und aufgrund der Symmetrie auch $L_6 := \{w \in \Sigma^* : |w|_b \leq |w|_a\}$. Dann wäre aber auch der Schnitt $L_5 \cap L_6$ eine reguläre Sprache. Wir haben aber bereits gezeigt, dass die Menge aller Wörter mit $|w|_a = |w|_b$ nicht regulär ist.

Alternativ kann man ein Pumpingargument benutzen: Angenommen L_1 wäre regulär. Dann sei n die Schranke für L_1 aus dem Pumping Lemma. Wir betrachten das Wort $x = a^{n+1}b^n$. Nach dem Lemma gibt es eine Zerlegung $x = uvw$ mit $|uv| \leq n, v \neq \epsilon$ und $uv^m w \in L_1$ für alle $m \in \mathbb{N}$. Da $|uv| \leq n$ und $v \neq \epsilon$, ist $v = a^k$ für $k \geq 1$. Also ist $uv^0 w = a^{n+1-k}b^n \notin L_1$. Widerspruch!

L_2: L_2 ist nicht kontextfrei. Angenommen L_2 wäre kontextfrei. Sei n die Konstante aus dem Pumping Lemma und setze $x = a^{n+2}b^{n+1}c^n$. Nach dem Lemma gibt es eine Zerlegung $x = yuvwz$ mit $|uvw| \leq n, uw \neq \epsilon$ und $yu^m vw^m z \in L_2$ für alle $m \in \mathbb{N}$. Da $|uvw| \leq n$ kann uvw nicht sowohl a als auch c enthalten. Deshalb gibt es nur die folgenden Möglichkeiten:

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

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

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

ein Durchschnitt von zwei regulären Sprachen und deshalb regulär. Es folgt, dass $L_3 = \bigcup_{n \leq 2010} L_3^n$ auch regulär ist.

L_4: L_4 ist kontextfrei, aber nicht regulär. $L_4 = L_2 \cap \{x \in \Sigma^* : |x|_b \geq 2010\}$ ist der Durchschnitt einer kontextfreien Sprache mit einer regulären Sprache und deshalb kontextfrei. Wäre L_4 auch noch regulär, dann wäre $L_1 = L_3 \cup L_4$ das auch. Wir haben aber oben gezeigt, dass L_1 nicht regulär ist.