Prof. Dr. Jasmin Blanchette
Jannis Limperg
Lydia Kondylidou

Ludwig-Maximilians-Universität München
Institut für Informatik
Abgabe bis 24. Mai 2023, 10:00 Uhr

Lösungsvorschlag zur Übung 5 zur Vorlesung
Formale Sprachen und Komplexität

FSK5-1 ***Pumping-Lemma*** (2 Punkte)

Geben Sie für die folgenden Sprachen über $\{a, b, c\}$ an, ob diese regulär oder nicht regulär sind, und beweisen Sie, dass Ihre Antwort stimmt.

Um zu beweisen, dass eine Sprache nicht regulär ist, verwenden Sie das Pumping-Lemma. Um zu beweisen, dass eine Sprache regulär ist, genügt es, ein Konstrukt K anzugeben, das die Sprache erzeugt/erkennt. Sie müssen nicht beweisen, dass $L(K) = L$ ist.

a) $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{N}\}$

LÖSUNGSVORSCHLAG: Ist regulär, da $L = L(a^* b^* c^*)$.

b) $L = \{a^i b^j \mid i, j \in \mathbb{N}\} \cup \{b^i c^j \mid i, j \in \mathbb{N}\}$

LÖSUNGSVORSCHLAG: Ist regulär, da $L = L(a^* b^* \mid b^* c^*)$.

c) $M = \{ww \mid w \in \{a, b\}^*\}$

LÖSUNGSVORSCHLAG:

Sei $n \geq 1$ die beliebige Pumpingzahl.

Wir wählen das Wort $z = a^n b a^n b$ mit $|z| \geq n$. Für alle Zerlegungen $z = uvw$ mit $|uv| \leq n$ und $|v| \geq 1$ gilt $u = a^k$, $v = a^l$, $w = a^m b a^n b$, $k + l + m = n$, $k + l \leq n$, $l > 0$.

Nun ist aber $uv^2w = a^k a^{2l} a^m b a^n b$ nicht mehr in M, da $k + 2l + m = n + l \neq n$ für $l > 0$.

Damit ist die Pumpingeigenschaft für M widerlegt und das Pumping-Lemma zeigt, dass M nicht regulär ist.

d) $N = \{a^{3k+2} \mid k \in \mathbb{N}, k \geq 2\}$

LÖSUNGSVORSCHLAG: Ist regulär, da $N = L(aaaaaaaa(aaa)^*)$.

e) $Q = \{a^{(k+2)!} \mid k \in \mathbb{N}\}$

LÖSUNGSVORSCHLAG:

Sei $n \geq 1$ die beliebige Pumpingzahl.

Wir wählen das Wort $w = a^{(n+2)!}$ mit $|w| \geq n$. Betrachte nun jede beliebige Zerlegung $z = uvw$ mit $|uv| \leq n$ und $|v| \geq 1$. Dann gilt $u = a^k, v = a^l, w = a^m$ mit $k + l + m = (n+2)!, k + l \leq n, l > 0$.

Dann ist $uv^iw = a^{k+il+m} = a^{(n+2)!+(i-1)l}$. Für $i = 2$ ergibt sich $uv^2w = a^{k+2l+m} = a^{(n+2)!+l}$. Da $l \leq n$, haben wir $(n+2)! < (n+2)! + l < ((n+1)+2)!$. Daher ist $a^{(n+2)!+l}$ nicht von der Form $a^{(k+2)!}$, d.h. $a^{(n+2)!+l} \notin Q$.

Damit ist die Pumpingeigenschaft für Q widerlegt und das Pumping-Lemma zeigt, dass Q nicht regulär ist.

f) $R = \{w \mid \#_a(w) = \#_b(w)\}$

LÖSUNGSVORSCHLAG:

Sei $n \geq 1$ die beliebige Pumpingzahl.

Wir wählen das Wort $w = a^n b^n$ mit $|w| \geq n$. Betrachte nun jede beliebige Zerlegung $z = uvw$ mit $|uv| \leq n$ und $|v| \geq 1$. Dann gilt $u = a^k, v = a^l, w = a^m b^n$ mit $k + l + m = n, k + l \leq n, l > 0$.

Dann ist $uv^iw = a^{k+il+m}b^n = a^{n+(i-1)l}b^n$. Für $i = 2$ ergibt sich $uv^2w = a^{n+l}b^n$. Da $l > 0$, haben wir $a^{n+l}b^n \notin R$.

Damit ist die Pumpingeigenschaft für R widerlegt und das Pumping-Lemma zeigt, dass R nicht regulär ist.

FSK5-2 ***Myhill und Nerode*** (2 Punkte)

a) Sei $L = L(ab^*cb^*a)$ über dem Alphabet $\Sigma = \{a, b, c\}$. Geben Sie eine kompakte Beschreibung der Äquivalenzklasse der Nerode-Relation von L an, in der das Wort $u = abc$ liegt.

LÖSUNGSVORSCHLAG:

Für $w \in \Sigma^*$ gilt $uw \in L$ g.d.w. $w \in L(b^*a)$: Für jedes solche w liegt uw in L; hat w eine andere Form, kann das Wort nicht mehr in L liegen. Dies ist die verbindende Eigenschaft all derjenigen Elemente, welche in der gleichen Äquivalenzklasse wie u liegen.

Umgekehrt kann man nun alle v bestimmen, für die vw mit $w \in L(b^*a)$ in der Sprache L liegt. Genau diese Wörter v sind äquivalent zu u bezüglich der Nerode-Relation, d.h. sie sind in der Äquivalenzklasse von u enthalten. Offensichtlich sind diese v genau die Wörter aus $L(ab^*cb^*)$, also $[u] = L(ab^*cb^*)$.

b) Sei $L = \{a^i b^j a^i \mid i, j \in \mathbb{N}_{>0}\}$ über dem Alphabet $\Sigma = \{a, b\}$. Geben Sie eine kompakte Beschreibung der Äquivalenzklasse der Nerode-Relation von L an, in der das Wort $u = aaabaa$ liegt.

LÖSUNGSVORSCHLAG:

Für $w \in \Sigma^*$ gilt $uw \in L$ g.d.w. $w = a$: Für ein solches w liegt uw in L; hat w eine andere Form, kann das Wort nicht mehr in L liegen. Dies ist die verbindende Eigenschaft all derjenigen Elemente, welche in der gleichen Äquivalenzklasse wie u liegen.

Umgekehrt kann man nun alle v bestimmen, für die vw mit $w = a$ in der Sprache L liegt. Genau diese Wörter v sind äquivalent zu u bezüglich der Nerode-Relation, d.h. sie sind in der Äquivalenzklasse von u enthalten. Offensichtlich sind diese v genau die Wörter aus $\{a^i b^j a^{i-1} \mid i, j \in \mathbb{N}_{>0}\}$, also $[u] = \{a^i b^j a^{i-1} \mid i, j \in \mathbb{N}_{>0}\}$.

c) Bestimmen Sie den Nerode-Index folgender Sprachen und entscheiden Sie mit dem Satz von Myhill und Nerode, welche dieser Sprachen regulär sind. Geben Sie für jede Sprache mit endlichem Nerode-Index alle paarweise disjunkten Äquivalenzklassen an (1 Repräsentant pro Klasse).

- $\{aaa, aab, aba, abb, baa, bab, bba, bbb\}$ mit $\Sigma = \{a, b\}$

LÖSUNGSVORSCHLAG:

Paarweise verschiedene Äquivalenzklassen: $[\varepsilon]$, $[a]$, $[aa]$, $[aaa]$, $[aaaa]$.

Damit ist der Nerode-Index 5 und die Sprache ist regulär. (Nebenbei: eine endliche Sprache kann ohnehin nur endlich viele Äquivalenzklassen haben, muss also regulär sein.)

- $\{a^{3n+2} \mid n \in \mathbb{N}_{>0}\}$ mit $\Sigma = \{a\}$

LÖSUNGSVORSCHLAG:
Paarweise verschiedene Äquivalenzklassen: $[\varepsilon], [a], [a^2], [a^3], [a^4], [a^5]$.
Damit ist der Nerode-Index 6 und die Sprache ist regulär.

- $\{w\overline{w} \mid w \in \Sigma^*\}$ mit $\Sigma = \{a, b\}$

LÖSUNGSVORSCHLAG:
Jedes Wort der Menge $M = \{a^i b \mid i \in \mathbb{N}\}$ liegt in einer anderen Äquivalenzklasse, denn für $u = ba^i$ gilt: $a^i bu \in L$, aber $a^j bu \notin L$ (für $j \neq i$). Somit ist $[a^i b] \neq [a^j b]$.
Da M unendlich viele Wörter enthält, gibt es auch unendlich viele Äquivalenzklassen. Das heißt der Nerode-Index ist ∞ und die Sprache ist nicht regulär.

- $\{w^i \mid i \in \mathbb{N}, w \in \Sigma^*\}$ mit $\Sigma = \{a, b\}$

LÖSUNGSVORSCHLAG:
Da $1 \in \mathbb{N}$, enthält die Sprache alle Wörter $w \in \Sigma^*$. Die einzige Aquivalenzklasse ist $[\varepsilon]$.
Damit ist der Nerode-Index 1 und die Sprache ist regulär.

FSK5-3 ***Konservative Erweiterungen regulärer Ausdrücke*** (0 Punkte)

In dieser Aufgabe geht es um Erweiterungen regulärer Ausdrücke mit zusätzlichen Operatoren. Geben Sie für jede Erweiterung an, ob sie konservativ ist, und beweisen Sie Ihre Antwort. Eine Erweiterung ist konservativ, wenn die erweiterten regulären Ausdrücke nur reguläre Sprachen beschreiben.

Im Folgenden seien $\alpha, \alpha_0, \alpha_1$ reguläre Ausdrücke.

a) $\alpha?$: Teilwörter, die von α erkannt werden, dürfen vorkommen, müssen aber nicht. Die Semantik von $\alpha?$ ist also $L(\alpha?) = \{\varepsilon\} \cup L(\alpha)$.

LÖSUNGSVORSCHLAG: Die Erweiterung ist konservativ. Es ist $L(\alpha?) = L(\alpha \mid \varepsilon)$, das heißt jeder Teilausdruck $\alpha?$ kann durch den regulären Ausdruck $\alpha \mid \varepsilon$ ersetzt werden, ohne die Bedeutung des gesamten Ausdrucks zu ändern.

b) α^+: wie α^*, aber α muss mindestens einmal vorkommen.

$L(\alpha^+) = \bigcup_{i \in \mathbb{N}_{>0}} L(\alpha)^i = L(\alpha) \cup L(\alpha)^2 \cup L(\alpha)^3 \cup \cdots$

LÖSUNGSVORSCHLAG: Konservativ, denn $L(\alpha^+) = L(\alpha\alpha^*)$.

c) $\alpha^{\{i,j\}}$ mit $i, j \in \mathbb{N}$ und $i \leq j$: wie α^*, aber α muss mindestens i-mal und darf höchstens j-mal wiederholt werden.

$L(\alpha^{\{i,j\}}) = \bigcup_{k=i}^{j} L(\alpha)^k = L(\alpha)^i \cup L(\alpha)^{i+1} \cup L(\alpha)^{i+2} \cup \cdots \cup L(\alpha)^j$

LÖSUNGSVORSCHLAG: Konservativ, denn $L(\alpha^{\{i,j\}}) = L(\underbrace{\alpha\alpha\ldots\alpha}_{i\text{-mal}}\underbrace{\alpha?\alpha?\ldots\alpha?}_{(j-i)\text{-mal}})$

d) $s(\alpha)$ („Shuffle-Operator“): die Buchstaben von α dürfen in beliebiger Reihenfolge auftauchen.

$L(s(\alpha)) = \{w \mid \exists v \in L(\alpha).\ \forall a \in \Sigma.\ \#_a(v) = \#_a(w)\}$

Beispiel: $L(s(aab)) = \{aab, aba, baa\}$.

LÖSUNGSVORSCHLAG: Nicht konservativ, das heißt Ausdrücke mit s können auch nicht-reguläre Sprachen erkennen.

Beispielsweise ist $s((ab)^*)$ nicht regulär. Es handelt sich hierbei um die Sprache von Wörtern über $\{a, b\}$, die gleich viele as und bs enthalten. Die Nicht-Regularität kann mit dem Pumping-Lemma gezeigt werden oder mit dem Satz von Myhill und Nerode.

Alternativ ist auch $L(s((ab)^*)) \cap L(a^*b^*) = \{a^ib^i \mid i \in \mathbb{N}\}$, was eines der Standardbeispiele für nicht reguläre Sprachen ist.

FSK5-4 ***Produktautomat*** (0 Punkte)

Als Produktautomat bezeichnet man einen DFA $C = (Z_C, \Sigma, \delta_C, z_{0C}, E_C)$, der aus zwei DFAs $A = (Z_A, \Sigma, \delta_A, z_{0A}, E_A)$ und $B = (Z_B, \Sigma, \delta_B, z_{0B}, E_B)$ gebildet wird, wobei

- $Z_C = Z_A \times Z_B$
- $z_{0C} = (z_{0A}, z_{0B})$

- $\delta_C((p,q),a) = (\delta_A(p,a), \delta_B(q,a))$.

Die Menge der Endzustände E_C werden wir in den folgenden Teilaufgaben unterschiedlich definieren.

a) Ein Automat X *simuliert* einen Automaten Y, wenn es eine Simulationsfunktion $f : Z_X \to Z_Y$ gibt, sodass für jedes Wort $w \in \Sigma^*$ gilt: Endet der Lauf von X auf w im Zustand z, so endet der Lauf von Y auf w im Zustand $f(z)$.

   Zeigen Sie, dass C sowohl A als auch B simuliert.

LÖSUNGSVORSCHLAG:

Die Simulationsfunktion für A ist $f_A(p,q) = p$, die für B $f_B(p,q) = q$.

Wir zeigen per Induktion über die Wortlänge n, dass diese Funktionen die gewünschte Eigenschaft haben.

- $n = 0$: Damit ist $w = \varepsilon$. Der letzte Zustand des Laufs eines DFA auf ε ist der Startzustand. Der Startzustand von C ist (z_{0A}, z_{0B}). Es ist $f_A(z_{0A}, z_{0B}) = z_{0A}$, der Startzustand von A und $f_B(z_{0A}, z_{0B}) = z_{0B}$, der Startzustand von B.
- $n \to n+1$:
  $|w| = n+1 > 0$, also gibt es $u \in \Sigma^*, a \in \Sigma$ mit $w = ua$.
  Seien z_{uA} der Zustand, in dem der Lauf von A auf u endet und z_{uB} der Zustand, in dem der Lauf von B auf u endet.
  Da $|u| = n$, endet der Lauf von C auf u im Zustand (z_{uA}, z_{uB}).
  Der Lauf von A auf $w = ua$ endet im Zustand $\delta_A(z_{uA}, a)$ und der Lauf von B auf w im Zustand $\delta_B(z_{uB}, a)$.
  Der Lauf von C endet im Zustand $\delta_C((z_{uA}, z_{uB}), a) = (\delta_A(z_{uA}, a), \delta_B(z_{uB}, a))$.
  Die geforderte Eigenschaft folgt, da $f_A(\delta_A(z_{uA}, a), \delta_B(z_{uB}, a)) = \delta_A(z_{uA}, a)$ sowie $f_B(\delta_A(z_{uA}, a), \delta_B(z_{uB}, a)) = \delta_B(z_{uB}, a)$.

b) Bestimmen Sie eine Menge I, sodass mit $E_C = I$ gilt: $L(C) = L(A)$.

LÖSUNGSVORSCHLAG:

$I = \{(p,q) \mid p \in E_A\}$.

$L(C) = L(A)$, da für jedes $w \in \Sigma^*$ gilt:

- w wird von A akzeptiert $\longleftrightarrow$ letzter Zustand im Lauf von A auf w endet in einem Zustand z mit $z \in E_A$

- w wird von C (mit Endzustandsmenge I) akzeptiert $\longleftrightarrow$ letzter Zustand im Lauf von C auf w endet in einem Zustand (z, y) mit $(z, y) \in I \longleftrightarrow$ letzter Zustand im Lauf von C auf w endet in einem Zustand (z, y) mit $z \in E_A$

Aufgrund der Simulationseigenschaft von f_A sind diese Bedingungen äquivalent, also $L(A) = L(C)$.

c) Bestimmen Sie eine Menge J, sodass mit $E_C = J$ gilt: $L(C) = L(A) \triangle L(B)$. Hierbei ist $M \triangle N = (M \cup N) \backslash (M \cap N)$ die „symmetrische Differenz" von M und N.

LÖSUNGSVORSCHLAG:

$J = \{(p, q) \mid p \in E_A \dot{\vee} q \in E_B\}$, wobei $\dot{\vee}$ das exklusive Oder ist. Es gilt also $A \dot{\vee} B \iff (A \vee B) \wedge \neg(A \wedge B)$.

$L(C) = L(A) \triangle L(B)$, da für jedes $w \in \Sigma^*$ gilt:

- w wird von A akzeptiert $\longleftrightarrow$ letzter Zustand im Lauf von A auf w endet in einem Zustand z mit $z \in E_A$
- w wird von B akzeptiert $\longleftrightarrow$ letzter Zustand im Lauf von B auf w endet in einem Zustand z mit $z \in E_B$
- w wird von C akzeptiert $\longleftrightarrow$ letzter Zustand im Lauf von C auf w endet in einem Zustand (z, y) mit $(z, y) \in J \longleftrightarrow$ letzter Zustand im Lauf von C auf w endet in einem Zustand (z, y) mit $z \in E_A \dot{\vee} y \in E_B$

Damit ist $L(C) = \{w \mid w \in L(A) \dot{\vee} w \in L(B)\} = L(A) \triangle L(B)$.