Prof. Dr. Jasmin Blanchette
Elisabeth Lempa
Luca Maio

Ludwig-Maximilians-Universität München
Institut für Informatik
Besprechung 30.05.2025 bis 02.06.2025
Abgabe bis 09.06.2025, 10:00 Uhr

Lösungsvorschlag zur Übung 5 zur Vorlesung
Formale Sprachen und Komplexität

FSK5-1 *Myhill und Nerode*

a) Sei $L = L(ab^*c)$ eine Sprache über dem Alphabet $\Sigma = \{a, b, c\}$. Geben Sie für jedes der folgenden Wörter u_i, $i \in \{1, 2, 3\}$, eine kompakte Beschreibung der Äquivalenzklasse $[u_i]$ der Nerode-Relation von L an.

i) $u_1 = abb$

> **LÖSUNGSVORSCHLAG:**
> Die Nerode-Äquivalenzklasse $[u_1]$ ist die Menge der Wörter v sodass für alle Wörter w gilt: $u_1w \in L$ g.d.w. $vw \in L$. Für $u_1 = abb$ sind die Wörter w mit $u_1w \in L$ genau die Wörter in der Sprache $L(b^*c)$. Die Wörter v mit $vw \in L$ für alle $w \in L(b^*c)$ sind genau die Wörter in der Sprache $L(ab^*)$. Somit ist $[abb] = L(ab^*)$.

ii) $u_2 = \varepsilon$

> **LÖSUNGSVORSCHLAG:**
> Es ist $\varepsilon w \in L$ g.d.w. $w \in L$. Für solche w gilt $vw \in L$ g.d.w. $v = \varepsilon$, denn für jedes andere v kann vw nicht genau ein a am Beginn des Wortes enthalten. Somit ist $[\varepsilon] = \{\varepsilon\}$.

iii) $u_3 = c$

> **LÖSUNGSVORSCHLAG:**
> Es gibt kein w sodass $cw \in L$ ist. Somit ist $[c]$ die Menge der Wörter $v \in \Sigma^*$, für die mit jedem w gilt: $vw \notin L$. Das ist der Fall für $v \in L(b\Sigma^*)$, $v \in L(c\Sigma^*)$, $v \in L(aa\Sigma^*)$, $v \in L(ab^*a\Sigma^*)$ und $v \in L(ab^*c\Sigma\Sigma^*)$, wobei Σ der reguläre Ausdruck $(a|b|c)$ ist. Somit ist
>
> $$[c] = L((b|c|aa)\Sigma^* \mid ab^*a\Sigma^* \mid ab^*c\Sigma\Sigma^*)$$

b) Bestimmen Sie den Nerode-Index folgender Sprachen L_i, $i \in \{1, 2, 3\}$, über Alphabeten Σ_i und entscheiden Sie mit dem Satz von Myhill und Nerode, welche der

Sprachen regulär sind. Geben Sie für jede Sprache mit endlichem Nerode-Index alle paarweise verschiedenen Äquivalenzklassen an (1 Repräsentant pro Klasse).

i) $L_1 = \{aaab, aabb, abab, abbb, baab, babb, bbab, bbbb\}$ mit $\Sigma_1 = \{a, b\}$

LÖSUNGSVORSCHLAG:
Paarweise verschiedene Äquivalenzklassen:

$$[\varepsilon], [a], [aa], [aaa], [aaab], [aaaa]$$

Damit ist der Nerode-Index 6 und die Sprache ist regulär. (Nebenbei: Eine endliche Sprache kann nur endlich viele Äquivalenzklassen haben, muss also regulär sein.)

ii) $L_3 = \{a^i b^j c^k \mid i, j, k \in \mathbb{N} \text{ und wenn } i = 2, \text{ dann } j < k\}$ mit $\Sigma_3 = \{a, b, c\}$

LÖSUNGSVORSCHLAG:
Der Nerode-Index ist unendlich. Betrachte für jedes $m \in \mathbb{N}$ die Wörter $u_m = aab^m$ und $w_m = c^{m+1}$. Es ist $u_m w_m = aab^m c^{m+1} \in L_3$, aber $u_n w_m = aab^n c^{m+1} \notin L_3$ für jedes $n > m$. Somit ist $[u_1] \neq [u_n]$ für jedes $n > 1$, $[u_2] \neq [u_n]$ für jedes $n > 2$, etc. Es gibt also unendlich viele verschiedene Äquivalenzklassen, d.h. $Index(\sim_{L_3}) = \infty$. Somit ist L_3 nicht regulär.

FSK5-2 ***Pumping-Lemma für kontextfreie Sprachen***

Zeigen Sie mit dem Pumping-Lemma für kontextfreie Sprachen, dass die Sprache $L = \{w\overline{w}w \mid w \in \Sigma^*\}$ über dem Alphabet $\Sigma = \{a, b\}$ nicht kontextfrei ist.

LÖSUNGSVORSCHLAG:
Beweis mit dem Pumping-Lemma.
Sei $n \in \mathbb{N}_{>0}$ beliebig.
Wir wählen $z \in L$ als $z = a^n b^{2n} a^{2n} b^n$ mit $|z| \geq n$.
Sei $z = uvwxy$ eine beliebige Zerlegung von z, sodass $|vx| \geq 1$, $|vwx| \leq n$ und $uv^i wx^i y \in L$ für jedes $i \in \mathbb{N}$. Wir wählen $i = 0$. Sei $z' = uv^0 wx^0 y$. Sei $k = |vx|$.
Wegen $|vwx| \leq n$ sind nur folgende Fälle möglich:

- $z' = a^{n-k_1} b^{2n-k_2} a^{2n} b^n$ mit $k_1 + k_2 = k \geq 1$. Da $z' \in L$ ist, gibt es ein w, sodass $z' = w\overline{w}w$ ist. Dieses w muss von der Form $a^* b^*$ sein. Somit muss sowohl $2(n - k_1) = 2n$ gelten als auch $2n - k_2 = 2n$. Diese Aussagen können aber nicht beide wahr sein. Widerspruch.

- $z' = a^n b^{2n-k_1} a^{2n-k_2} b^n$ mit $k_1 + k_2 = k \geq 1$. Dann ist (mit der gleichen

Begründung wie oben) $z' \in L$ nur wenn sowohl $2n = 2n - k_2$ als auch $2n = 2n - k_1$ gilt, aber diese Aussagen können nicht beide wahr sein. Widerspruch.

- $z' = a^n b^{2n} a^{2n-k_1} b^{n-k_2}$ mit $k_1 + k_2 = k \geq 1$. Dann ist $z' \in L$ nur wenn sowohl $2n = 2n - k_1$ als auch $2n = 2(n - k_2)$ gilt, aber diese Aussagen können nicht beide wahr sein. Widerspruch.

FSK5-3 ***Reguläre und nicht-reguläre Sprachen***

a) Zeigen Sie, dass die Sprache $\{a^i b^j c^k \mid i, j, k \in \mathbb{N}$ und wenn $i = 2$, dann $j < k\}$ über dem Alphabet $\Sigma = \{a, b, c\}$ die Pumping-Eigenschaft erfüllt.

LÖSUNGSVORSCHLAG:
Wir wählen $n = 4$.
Sei $z = a^i b^j c^k \in L$ mit $|z| \geq n$.
Wir müssen für jedes solche z eine Zerlegung $z = uvw$ angeben mit $|uv| \leq n$, $|v| \geq 1$ und $uv^l w \in L$ für alle $l \in \mathbb{N}$.
Vollständige Fallunterscheidung:

- $i = 2$. Wähle die Zerlegung $u = \varepsilon$, $v = aa$, $w = b^j c^k$. Es ist $uv^l w \in L$ für jedes $l \in \mathbb{N}$, denn:
  - Für $l = 1$ ist $uv^l w = z \in L$.
  - Für $l \neq 1$ enthält $uv^l w$ nicht genau 2 a's und somit ist die Anzahl der b's und c's wieder irrelevant.
- $i \geq 4$. Wähle die Zerlegung $u = a^3$, $v = a$, $w = a^{i-4} b^j c^k$ mit $|uv| \leq 4$ und $|v| \geq 1$. Es ist $uv^l w \in L$ für jedes $l \in \mathbb{N}$, da die Anzahl der a's in $uv^l w$ immer mindestens 3 ist und somit die Anzahl der b's und c's irrelevant ist.
- $i < 4$ und $i \neq 2$. Wegen $|z| \geq 4$ ist $j \geq 1$ oder $k \geq 1$.
  - Wenn $j \geq 1$: Wähle die Zerlegung $u = a^i$, $v = b$, $w = b^{j-1} c^k$ mit $|uv| \leq 4$ und $|v| \geq 1$. Es ist $uv^l w \in L$ für jedes $l \in \mathbb{N}$, da die Anzahl der a's in $uv^l w$ immer $i \neq 2$ bleibt und somit die Anzahl der b's und c's irrelevant ist.
  - Wenn $j = 0$: Wähle die Zerlegung $u = a^i$, $v = c$, $w = c^{k-1}$. Diese Zerlegung erfüllt mit analoger Begründung alle Bedingungen.

b) Sind die folgenden Sprachen L_i, $i \in \{1, 2, 3\}$, über den Alphabeten Σ_i regulär? Wenn ja, geben Sie einen regulären Ausdruck an, der L_i erkennt. (Sie müssen nicht beweisen, dass der reguläre Ausdruck L_i erkennt.) Wenn nein, zeigen Sie die Nichtregularität mit dem Pumping-Lemma für reguläre Sprachen.

i) $L_1 = \{ac^iba^jb \mid i, j \in \mathbb{N}\}$ mit $\Sigma_1 = \{a, b, c\}$

> **LÖSUNGSVORSCHLAG:**
> Regulär; $L_1 = L(ac^*ba^*b)$.

ii) $L_2 = \{a^pb^p \mid p \in \mathbb{N}$ ist prim$\}$ mit $\Sigma_2 = \{a, b\}$

> **LÖSUNGSVORSCHLAG:**
> Nicht regulär. Beweis mit dem Pumping-Lemma.
> Sei $n \in \mathbb{N}_{>0}$ beliebig.
> Wir wählen $z \in L_2$ als $z = a^pb^p$, wobei p die kleinste Primzahl mit $p \geq n$ ist. Damit ist $|z| \geq n$.
> Sei $z = uvw$ eine beliebige Zerlegung von z, sodass $|uv| \leq n$, $|v| \geq 1$ und $uv^iw \in L_2$ für jedes $i \in \mathbb{N}$. Da $|uv| \leq n$ ist, ist $v = a^k$ für ein $k \in \mathbb{N}_{>0}$.
> Wir wählen $i = 0$. Das Wort uv^iw hat weniger a's als b's und somit ist $uv^iw \notin L_2$. Widerspruch.

iii) $L_3 = \{a^{2n+1} \mid n \in \mathbb{N}\}$ mit $\Sigma_3 = \{a\}$.

> **LÖSUNGSVORSCHLAG:**
> Regulär; $L_3 = L(a(aa)^*)$.

FSK5-4 ***Konservative Erweiterungen regulärer Ausdrücke***

In der Praxis werden reguläre Ausdrücke häufig mit weiteren Operatoren erweitert. Eine solche Erweiterung ist *konservativ*, wenn die erweiterten regulären Ausdrücke nur reguläre Sprachen beschreiben. Geben Sie in jeder Teilaufgabe an, ob die beschriebene Erweiterung konservativ ist, und beweisen Sie Ihre Antwort. Dabei sei α ein regulärer Ausdruck über einem beliebigen Alphabet.

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 $L(\alpha?) = L(\alpha|\varepsilon)$, das heißt jeder Teilausdruck $\alpha?$ kann durch den regulären Ausdruck $\alpha|\varepsilon$ 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) $\backslash n$ mit $n \in \mathbb{N}$. In einem regulären Ausdruck α bezeichnen wir den n-ten Teilausdruck der Form (α_0) (wobei α_0 ein regulärer Ausdruck ist) als die n-te Capturing Group. Ein Teilausdruck $\backslash n$ in α wird dann als Backreference bezeichnet und erkennt genau die Zeichenkette, die von α_0 erkannt wurde. Beispielsweise erkennt $(a|b)\backslash 1$ die Wörter aa und bb, aber nicht ab oder ba.

LÖSUNGSVORSCHLAG:
Nicht konservativ. Der Ausdruck $((a|b)^*)\backslash 1$ erkennt die Sprache $\{ww \mid w \in \{a, b\}^*\}$, die bekanntlich nicht regulär ist.