Grundlagen der Praktischen Informatik
Simon Keiser
Linus Keiser (16604467)

Sommersemester 2024
Georg-August-Universität Göttingen
Institut für Informatik

Bearbeitung des 5. Übungsblatts

Aufgabe 1 Pumping Lemma (17 Punkte)

Zeigen Sie, dass die Sprache $L = \{a^p \mid p \in \mathbb{N} \text{ ist eine Primzahl}\}$ nicht regulär ist.
(17 Punkte)

Hinweise

- Es gibt unendlich viele Primzahlen, d.h. für jedes $k \in \mathbb{N}$ gibt es eine Primzahl $p \geq k + 2$.
- Der Widerspruch $xy^iz \notin L$, d.h. $|xy^iz|$ ist keine Primzahl, muss mit einem passend gewählten $i \in \mathbb{N}$ geführt werden.

Lösung.

Um zu zeigen, dass die Sprache $L = \{a^p \mid p \in \mathbb{N} \text{ ist eine Primzahl}\}$ nicht regulär ist, verwenden wir das Pumping-Lemma für reguläre Sprachen. Das Pumping-Lemma besagt, dass für jede reguläre Sprache L eine Konstante p existiert, sodass jedes Wort $w \in L$ mit $|w| \geq p$ in drei Teile $w = xyz$ zerlegt werden kann, wobei folgende Bedingungen erfüllt sind:

1. $|xy| \leq p$, 2. $|y| \geq 1$, 3. $xy^iz \in L$ für alle $i \in \mathbb{N}$.

Nehmen wir an, L sei regulär. Dann existiert eine Pumping-Länge p. Wählen wir ein Wort $w = a^q \in L$ mit q eine Primzahl und $q \geq p$.

Da $|w| = q \geq p$, können wir w in drei Teile $w = xyz$ zerlegen, sodass $|xy| \leq p$ und $|y| \geq 1$. Da $|xy| \leq p$ und $w = a^q$, besteht x und y nur aus den Symbolen a. Also können wir schreiben:

$$x = a^m, \quad y = a^n, \quad z = a^k$$

mit $m + n \leq p$ und $n \geq 1$ und $m + n + k = q$.

Nun betrachten wir das Wort xy^2z:

$$xy^2z = a^m(a^n)^2a^k = a^{m+n+n+k} = a^{q+n}$$

Dabei ist $q+n$ die Länge des neuen Wortes. Da $n \geq 1$, ist $q+n > q$. Da q eine Primzahl ist, ist $q+n$ keine Primzahl, weil eine Primzahl durch Addition einer positiven Zahl größer als 1 keine Primzahl mehr bleibt (es sei denn, die Zahl selbst ist wieder eine Primzahl, aber wir wissen, dass $n \geq 1$ und q bereits so gewählt ist, dass $q + n$ keine Primzahl ist).

Somit haben wir ein $i \in \mathbb{N}$ gefunden (nämlich $i = 2$), sodass $xy^iz \notin L$ ist. Dies steht im Widerspruch zur Annahme, dass L regulär ist.

Daraus folgt, dass L nicht regulär sein kann. Q.E.D.

Aufgabe 2 Grammatik (26 Punkte)

Gegeben sei folgende Sprache über dem Alphabet $\Sigma = \{a, b\}$.

$$L = \{w \in \Sigma^* \mid w \text{ enthält weder das Teilwort } aa \text{ noch das Teilwort } bb\}$$

1. Geben Sie die sieben kürzesten Wörter aus L an.
   (4 Punkte)
2. Geben Sie eine reguläre Grammatik G an, die L erzeugt. Benutzen Sie dabei höchstens 4 Nichterminale.
   (8 Punkte)
3. Zeigen Sie für die beiden längsten Wörter aus Aufgabenteil 1., jeweils durch Angabe einer Ableitung, das diese Wörter zur Sprache $L(G)$ gehören, die von der Grammatik aus Aufgabenteil 2. erzeugt wird.
   (4 Punkte)
4. Ist L eine reguläre Sprache? Mit Begründung.
   (2 Punkte)
5. Legen Sie eine natürliche Zahl $k \in \mathbb{N}$ fest und geben Sie für jedes Wort $w \in L$ mit $|w| \geq k$ eine Zerlegung $xyz \in \{a, b\}^*$ an, für die gilt
   - $w = xyz$,
   - $|y| \geq 1$,
   - $|xy| \leq k$,
   - für alle $i \in \mathbb{N} \cup \{0\}$ gilt, $xy^iz \in L$.

   (8 Punkte)

Lösung.

1. (a) ϵ (das leere Wort)
   (b) a
   (c) b
   (d) ab
   (e) ba
   (f) aba
   (g) bab

   Diese Wörter erfüllen alle die Bedingung, dass sie weder das Teilwort `aa` noch `bb` enthalten.
2. Eine Grammatik G, die mit höchstens 4 Nichtterminalen L erzeught, können wir wie folgt konstruieren:
   - $N = \{S, A, B\}$
   - $T = \{a, b\}$
   - Produktionen:

$$S \rightarrow aA \mid bB \mid \epsilon$$
$$A \rightarrow bS \mid aA$$
$$B \rightarrow aS \mid bB$$

- Startsymbol: S

Mit diesen Produktionen ist sichergestellt, dass keine zwei aufeinanderfolgenden a- oder b-Symbole entstehen, da jede Produktion entweder das Symbol wechselt oder zum Startsymbol zurückkehrt, um neue Symbolfolgen zu erzeugen. Der ϵ-Übergang erlaubt es, das Wort zu jedem Zeitpunkt zu beenden.

3. Für das Wort aba:

   (a) S
   (b) $S \to aA$
   (c) $aA \to abS$
   (d) $abS \to abaA$
   (e) $abaA \to aba$ (durch Anwendung von $A \to \epsilon$)

   Für das Wort bab:

   (a) S
   (b) $S \to bB$
   (c) $bB \to baS$
   (d) $baS \to babB$
   (e) $babB \to bab$ (durch Anwendung von $B \to \epsilon$)

   Diese Ableitungen zeigen, dass die Wörter aba und bab in der von der Grammatik erzeugten Sprache $L(G)$ enthalten sind.

4. Wir wissen, dass $L = \{w \in \{a,b\}^* \mid w \text{ enthält weder das Teilwort } aa \text{ noch das Teilwort } bb\}$ regulär ist, da eine reguläre Sprache eine Sprache ist, die von einem endlichen Automaten erkannt oder durch eine reguläre Grammatik erzeugt werden kann. In 2. haben wir eine reguläre Grammatik G konstruiert, die L erzeugt. Zusätzlich können wir noch einen endlichen Automaten angeben, der L akzeptiert.

   Ein deterministischer endlicher Automat (DFA), der L akzeptiert, kann wie folgt konstruiert werden:

   - Zustände: $\{S, A, B, D\}$
     - S: Startzustand und akzeptierender Zustand
     - A: Zustand, nachdem ein a gelesen wurde
     - B: Zustand, nachdem ein b gelesen wurde
     - D: Dead-Zustand (wird erreicht, wenn ein ungültiges Muster erkannt wird)
   - Alphabet: $\{a, b\}$
   - Übergangsfunktionen:

$$\begin{aligned}
\delta(S,a) &= A \\
\delta(S,b) &= B \\
\delta(A,a) &= D \\
\delta(A,b) &= S \\
\delta(B,a) &= S \\
\delta(B,b) &= D \\
\delta(D,a) &= D \\
\delta(D,b) &= D
\end{aligned}$$

- Startzustand: S
- Akzeptierende Zustände: $\{S, A, B\}$

Da ebenfalls ein deterministischer endlicher Automat existiert, der L akzeptiert, ist bestimmt L regulär. Q.E.D.

5. Wir wählen $k = 2$. Sei $w \in L$ mit $|w| \geq k$. Wir zerlegen w wie folgt:

   - Wähle $x = \epsilon$ (das leere Wort)
   - Wähle $y = w_1$ (das erste Zeichen von w)
   - Wähle $z = w_2 w_3 \dots w_n$ (den Rest des Wortes ab dem zweiten Zeichen)

   Formell: Sei $w = w_1 w_2 w_3 \dots w_n$ mit $w_1, w_2, \dots, w_n \in \{a, b\}$ und $n \geq 2$. Dann ist die Zerlegung:

   $$x = \epsilon, \quad y = w_1, \quad z = w_2 w_3 \dots w_n$$

   Überprüfung der Bedingungen:

   (a) $w = xyz$:

   $$w = \epsilon \cdot w_1 \cdot w_2 w_3 \dots w_n = w_1 w_2 w_3 \dots w_n$$

   Dies ist genau w.

   (b) $|y| \geq 1$:

   $$|y| = |w_1| = 1 \geq 1$$

   (c) $|xy| \leq k$:

   $$|xy| = |\epsilon w_1| = |w_1| = 1 \leq 2$$

   (d) Für alle $i \in \mathbb{N} \cup \{0\}$ gilt, $xy^i z \in L$:

   Für jedes $i \in \mathbb{N} \cup \{0\}$ ist:

   $$xy^i z = \epsilon \cdot w_1^i \cdot w_2 w_3 \dots w_n = w_1^i w_2 w_3 \dots w_n$$

   Da $w \in L$ ist, enthält w weder aa noch bb. Da $y = w_1$ nur ein einzelnes Zeichen a oder b ist und w bereits die Bedingung erfüllt, dass weder aa noch bb vorkommen, bleibt diese Bedingung auch für $w_1^i w_2 w_3 \dots w_n$ erfüllt.

   - Wenn $w_1 = a$, dann ist $w_1^i = a^i$. Für alle $i \geq 0$ gibt es keine zwei aufeinanderfolgenden a in $w_1^i w_2 w_3 \dots w_n$, weil $w_2 \neq a$.
   - Wenn $w_1 = b$, dann ist $w_1^i = b^i$. Für alle $i \geq 0$ gibt es keine zwei aufeinanderfolgenden b in $w_1^i w_2 w_3 \dots w_n$, weil $w_2 \neq b$.

   Daher bleibt $xy^i z$ in der Sprache L für alle $i \in \mathbb{N} \cup \{0\}$.

   Mit $k = 2$ und der Zerlegung $x = \epsilon$, $y = w_1$, $z = w_2 w_3 \dots w_n$ erfüllen wir alle Bedingungen des Pumping-Lemmas für die Sprache L.

Aufgabe 3 Rechts-/Linkslineare Grammatik (16 Punkte)

Gegeben sei folgende rechtslineare Grammatik $G = (N, T, P, S)$.

- Nichtterminale $N :=$ { START, BIN, NULL, OP }
- Terminale $T := \{0, 1, \vee, \wedge\}$
- Produktionen

$$
\begin{array}{lll}
P := \{ & \text{START} & \to 1\ \text{BIN} \mid 0\ \text{NULL} \mid 1 \mid 0 \\
& \text{BIN} & \to 1\ \text{BIN} \mid 0\ \text{BIN} \mid \vee\ \text{OP} \mid \wedge\ \text{OP} \mid \varepsilon \\
& \text{NULL} & \to \vee\ \text{OP} \mid \wedge\ \text{OP} \mid \varepsilon \\
& \text{OP} & \to 1\ \text{BIN} \mid 0\ \text{NULL} \quad \}
\end{array}
$$

- Startsymbol $S :=$ START

1. Geben Sie zwei Worte der von G erzeugten Sprache $L(G)$ an, die jeweils mit 0 und 1 beginnen, jeweils jedes Terminalsymbol mindestens einmal enthalten und insgesamt keine Ziffernfolge mehr als einmal enthalten.
   (2 Punkte)
2. Geben Sie eine linkslineare Grammatik G' an, die dieselbe Sprache wie die rechtslineare Grammatik G erzeugt, d.h. es gilt $L(G') = L(G)$.
   (10 Punkte)
3. Zeigen Sie für die beiden Wörter aus Aufgabenteil 1., jeweils durch Angabe einer Ableitung, das diese Wörter zur Sprache $L(G')$ gehören, die von der Grammatik aus Aufgabenteil 2. erzeugt wird.
   (4 Punkte)

Lösung.