Aufgabe 4.1 [Pumping-Lemma] **6 Punkte**

Zeigen Sie mit Hilfe des Pumping-Lemmas, dass die folgenden Sprachen über dem Alphabet $\{a, b, c\}$ nicht regulär sind.

a) $L_1 = \{cucv \mid u, v \in \{a, b\}^*, |u| = |v|\}$ **(3 Punkte)**

b) $L_2 = \{ua^i \mid u \in \{a, b, c\}^*, i \in \mathbb{N}_0, \#_a(u) = i\}$ **(3 Punkte)**

Lösung:

Um mit Hilfe des Pumping-Lemmas zu zeigen, dass die gegebenen Sprachen nicht regulär sind, können wir das Pumping-Lemma selbst verwenden oder aber Korollar 5.6, welches die Kontraposition des Pumping-Lemmas ist. In Aufgabenteil a) verwenden wir das Pumping-Lemma und in Aufgabenteil b) Korollar 5.6, um Beispiele für beide Ansätze zu bieten. Beide Beweise lassen sich aber auch problemlos mit dem jeweils anderen Ansatz führen.

a) Wir führen einen Beweis durch Widerspruch. Angenommen L_1 ist regulär. Dann können wir das Pumping-Lemma anwenden (und einen Widerspruch ableiten).

Sei $n \in \mathbb{N}$ gemäß dem Pumping-Lemma für L_1. Wir wählen außerdem das Wort $w = ca^nca^n$. Es gilt $w \in L_1$ (wähle $u = v = a^n$, damit gilt offensichtlich $|u| = |v|$) und $|w| = 2(n + 1) \geq n$.

Nach dem Pumping-Lemma gibt es eine Zerlegung $w = xyz$ mit

(1) $y \neq \varepsilon$,

(2) $|xy| \leq n$, und

(3) für alle $k \geq 0$ ist $xy^kz \in L_1$.

Wir zeigen nun, dass es ein $k \geq 0$ gibt, sodass $xy^kz \notin L_1$ gilt – ein Widerspruch zu (3) und damit zu der Annahme, dass L_1 regulär ist. Wegen (1) und (2) gibt es ein $j \leq n$, sodass $xy = ca^j$ und $z = a^{n-j}ca^n$ ist. Wir machen eine Fallunterscheidung nach y:

1. Fall: Das Teilwort y enthält ein c. Es folgt, dass $x = \varepsilon$ ist und $y = ca^j$. Wir wählen $k = 0$. Dann ist $xy^kz = xz = \varepsilon \cdot a^{n-j}ca^n$ und, weil xz in diesem Fall keine zwei c-Symbole enthält, $xy^kz \notin L_1$.

2. Fall: Das Teilwort y enthält kein c. Dann gibt es ein $i \leq j$, sodass $x = ca^{j-i}$ und $y = a^i$ ist. Wegen (1) gilt außerdem $i > 0$. Wir wählen $k = 2$. Dann ist $xy^kz = xyyz = ca^{j-i} \cdot a^i \cdot a^i \cdot a^{n-j}ca^n = ca^{n+i}ca^n$. Es gilt $|a^{n+i}| \neq |a^n|$, weil $i > 0$ ist, und damit $xy^kz \notin L_1$ (die Position der c-Symbole bestimmt u und v in der Sprachdefinition eindeutig).

In beiden Fällen erhalten wir einen Widerspruch zu (3) und damit zu der Annahme, dass L_1 regulär ist. Wir haben damit gezeigt, dass L_1 nicht regulär ist. □

Hinweis

In der Fallunterscheidung kann in beiden Fällen $k = 0$ (oder $k = 2$) gewählt werden. Hier haben wir bewusst verschiedene k gewählt, um die beiden Möglichkeiten $k = 0$ und $k = 2$ zu demonstrieren.

b) Wir nutzen Korollar 5.6 des Pumping-Lemmas, um zu zeigen, dass L_2 nicht regulär ist. Sei $n > 0$ beliebig. Wir wählen das Wort $w = a^n b a^n$. Es gilt $|w| = 2n + 1 \geq n$ und $w \in L_2$, denn wir können w als ua^n mit $u = a^n b$ schreiben, wobei $\#_a(u) = n$ gilt.

Wir zeigen nun, dass für jede Zerlegung $w = xyz$ mit

(1) $y \neq \varepsilon$ und

(2) $|xy| \leq n$

ein $k \geq 0$ existiert, sodass $xy^k z \notin L_2$ gilt. Dann folgt mit Korollar 5.6, dass L_2 nicht regulär ist.

Sei nun $w = xyz$ eine beliebige feste Zerlegung, die (1) und (2) erfüllt. Wegen (1) und (2) gibt es m und j mit $m \geq 0$, $j \geq 1$ und $m + j \leq n$, sodass $x = a^m$, $y = a^j$ und $z = a^{n-m-j} b a^n$ ist.

Wir wählen $k = 2$. Dann ist $xy^k z = xy^2 z = a^m \cdot a^{2j} \cdot a^{n-m-j} b a^n = a^{n+j} b a^n$. Wir zeigen als nächstes, dass $xy^2 z \notin L$ gilt. Für jede Zerlegung der Form va^i von $xy^2 z$ gilt, dass $v = a^{n+j} b a^\ell$ und $i = n - \ell$ für ein $\ell \geq 0$ sind, da das b-Symbol in v enthalten sein muss. Insbesondere gilt $\#_a(v) \geq n + j$ mit $j > 0$ und $i \leq n$. Das Wort $xy^2 z$ kann also nicht als va^i mit $\#_a(v) = i$ geschrieben werden. Es folgt $xy^2 z \notin L_2$ und damit, dass L_2 nicht regulär ist.

Hinweis

In dieser Beispiellösung kann *nicht* $k = 0$ gewählt werden. Wenn, beispielsweise, $n \geq 2$ und $y = 2$ ist, dann kann xz als va^i mit $i = n-1$ und $v = a^{n-2} ba$ geschrieben werden. Es gilt dann $xy^0 z = xz \in L_2$ und Korollar 5.6 kann nicht angewendet werden.

Aufgabe 4.2 [Zeichenkettensuche] 5 Punkte

Wir betrachten die Problemstellung MultiSearch aus dem Vorlesungskapitel 6. Ein Text sei gegeben als ein Wort über dem Alphabet $\{a, \ldots, z\}$. Es soll überprüft werden, ob eine der Zeichenketten $minimal$, $minimieren$ oder $maximal$ im Text enthalten ist.

a) Geben Sie einen NFA an, der genau dann akzeptiert, wenn eines der gesuchten Wörter im Text enthalten ist. Bezeichnen Sie dabei die Zustände des Automaten wie im Abschnitt 6.1.2 der Vorlesung („NFA zur Zeichenkettensuche: allgemein") mit dem bisher gelesenen Präfix. **(2 Punkte)**

b) Leiten Sie anschließend einen äquivalenten DFA wie in der Vorlesung beschrieben her. Die Herleitung soll direkt, wie in Abschnitt 6.1.2 der Vorlesung, erfolgen und *nicht* über die Potenzmengen-Konstruktion. Begründen Sie für mindestens eine zusätzlich eingefügte Transition die Korrektheit. **(3 Punkte)**

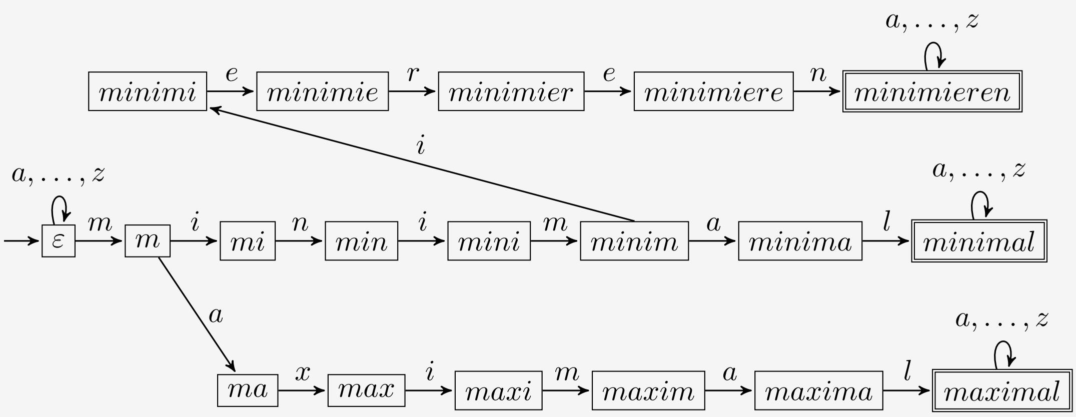
Hinweis

Der Übersicht halber müssen nicht alle Transitionen eingezeichnet werden. Nicht eingezeichnete Transitionen, die für das gleiche gelesene Zeichen in den gleichen Zustand führen, können unter Angabe einer geeigneten Transitionsfunktion $\delta(q, \sigma) = q_{pre}$, wobei q_{pre} dem Präfix des entsprechenden (Ziel-)Zustands entspricht, zusammengefasst werden.

Lösung:

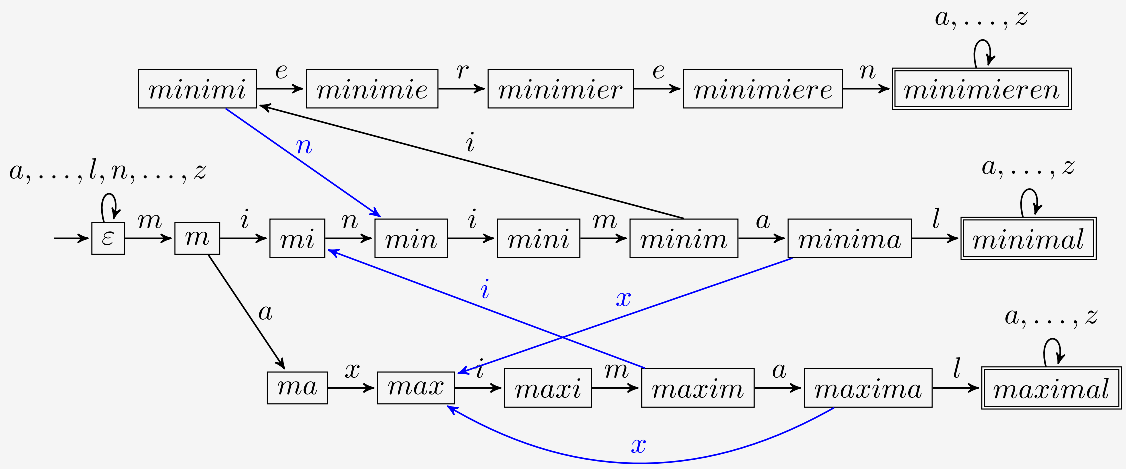
a) Für die drei Zeichenketten *minimal*, *minimieren*, und *maximal* erhalten wir mit dem Konstruktionsrezept aus der Vorlesung den folgenden NFA.

Für das Eingabewort $w = miminimalistisch$ kann so beispielsweise das Präfix $w_1 = mi$ gelesen werden, während der Automat in Zustand ε verbleibt (insbesondere wurde beim Lesen des ersten m die Transition $(\varepsilon, m, \varepsilon)$ verwendet). Beim Lesen das zweiten m kann der Automat dann in Zustand m übergehen (vermöge der Transition (ε, m, m)). Jedes weitere Zeichen des fixierten Suchwortes führt dann in den Zustand, der das nächstgrößere Präfix repräsentiert: $(m, i, mi), (mi, n, min), \ldots, (minima, l, minimal)$. Durch vollständiges Lesen eines der gesuchten Wörter kann der Automat den zugehörigen akzeptierenden Zustand erreichen (im Beispiel also den Zustand *minimal*). Durch die Transitionen $(minimal, \sigma, minimal)$ für jedes Zeichen σ akzeptiert der Automat dann unabhängig vom Rest der Eingabe (im Beispiel *istisch*).



b) Der geforderte DFA hat nach Vorlesung die gleichen Zustände, den gleichen Startzustand und dieselben akzeptierenden Zustände wie der NFA aus Teilaufgabe a).

Allerdings müssen Transitionen ergänzt, und im Falle des Startzustands entfernt werden. So wird beispielsweise eine Transition für den Zustand *minima* und das Zeichen x ergänzt, die in den Zustand *max* führt, weil *max* das längste Suffix von *minimax* ist, das Präfix eines gesuchten Wortes ist (hier *maximal*). Damit ist die beschriebene Transition nach Vorlesung korrekt.



Neben den blau markierten Transitionen wird außerdem $\delta(q, m) = m$ und $\delta(q, \sigma) = \varepsilon$ für $\sigma \neq m$ gesetzt, falls keine entsprechende Transition in der obigen Abbildung eingezeichnet ist.

Aufgabe 4.3 [Kontextfreie Grammatiken] — 4 Punkte

Sei G die folgende kontextfreie Grammatik.

$$
\begin{aligned}
S &\to B \mid C \\
B &\to aBa \mid bBb \mid D \\
C &\to cCa \mid cCb \mid D \\
D &\to Dc \mid \varepsilon
\end{aligned}
$$

a) Beschreiben Sie die Sprache $L(G)$ möglichst informell. Begründen Sie anschließend, warum Ihre Beschreibung korrekt ist, indem Sie auf die Bedeutung der einzelnen Variablen eingehen. **(1,5 Punkte)**

b) Geben Sie einen Ableitungsbaum und eine Ableitung für das Wort $w = aabaccabaa$ über G an. Es sind keine Begründungen erforderlich. **(1,5 Punkte)**

c) Ist die Grammatik G mehrdeutig? **(1 Punkt)**

Lösung:

a) Die Sprache $L(G)$ ist die Sprache aller Wörter,

- die mit einem Teilwort v über $\{a, b\}$ beginnen, gefolgt von einer beliebig langen Folge von c-Symbolen, und, unmittelbar nach der Folge von c-Symbolen, mit dem Spiegelwort v^R enden;

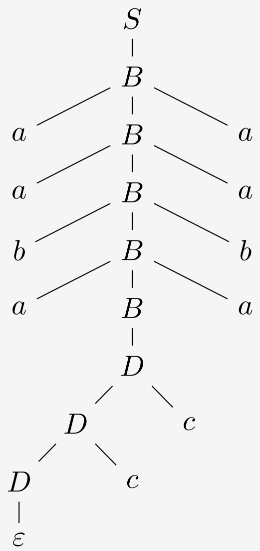
  oder

- die mit einer Folge von c-Symbolen beginnen und, unmittelbar darauf folgend, mit einem Teilwort über $\{a, b\}$ enden, das maximal so lang ist, wie die Folge von c-

Symbolen.

Die obige Fallunterscheidung lässt sich anhand der Regeln für das Startsymbol S festmachen. Die Sprache $L(G)$ ist die Sprache aller Wörter, die von der Variablen B (entspricht dem ersten Fall der informellen Beschreibung) *oder* von der Variablen C (entspricht dem zweiten Fall der informellen Beschreibung) ableitbar sind. Die Variable B wird dabei genutzt, um Satzformen der Gestalt vDv^R abzuleiten. Analog lassen sich mit Hilfe der Variablen C Satzformen der Gestalt c^iDv mit $|v| = i$ ableiten. Mit der Variablen D können zuletzt noch beliebig viele (weitere) c-Symbole erzeugt werden.

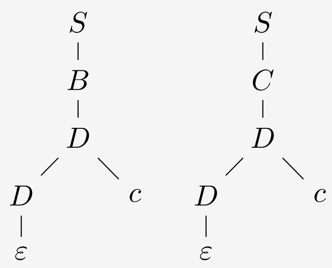
b) Ein Ableitungsbaum für w ist im Folgenden abgebildet.



Aus dem obigen Ableitungsbaum lässt sich folgende Ableitung für w ablesen. Die rechten Seiten, mit denen in einem Schritt die jeweils einzige Variable ersetzt wurde, sind hervorgehoben.

$$\begin{aligned}
S &\Rightarrow B \\
&\Rightarrow aBa \\
&\Rightarrow aaBaa \\
&\Rightarrow aabBbaa \\
&\Rightarrow aabaBabaa \\
&\Rightarrow aabaDabaa \\
&\Rightarrow aabaDcabaa \\
&\Rightarrow aabaDccabaa \\
&\Rightarrow aabaccabaa = w
\end{aligned}$$

c) Die Grammatik G ist mehrdeutig, weil es ein Wort gibt, für das es mindestens zwei verschiedene Ableitungsbäume gibt; und zwar hat beispielsweise das Wort c die folgenden zwei Ableitungsbäume.


S
B
D
D
c
ε
S
C
D
D
c
ε