

Übung 4 zur Vorlesung Formale Sprachen und Komplexität

FSK4-1 Reguläre Ausdrücke

(2 Punkte)

In dieser Aufgabe sind reguläre Ausdrücke entsprechend der Definition und Syntax im Vorlesungsskript, Definition 4.7.1, anzugeben.

- a) Geben Sie einen regulären Ausdruck an, der genau die Sprache

$$L = \{w \mid i \in \Sigma, w \in \Sigma^* \text{ und } \#_i(w) = i\}$$

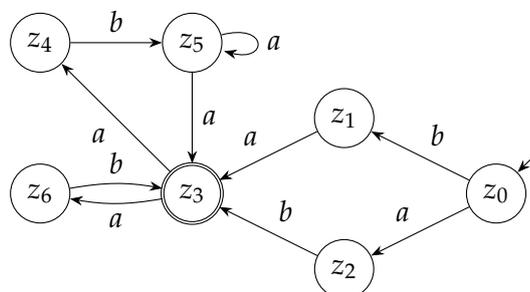
erzeugt, wobei $\Sigma = \{1, 2, 3\}$. (Siehe auch FSK3-1b.)

- b) Für Autokennzeichen in München gibt es folgende Regeln:

- Sie beginnen mit einem M , danach folgt ein Buchstabenfeld, danach folgt ein Ziffernfeld. Anschließend können noch ein H für Oldtimer oder ein E für Elektroautos kommen, aber nicht beides gleichzeitig.
- Das Buchstabenfeld besteht aus 1 oder 2 Buchstaben aus $\{A, \dots, Z\}$.
- Das Ziffernfeld besteht aus 3-4 Ziffern $\{0, \dots, 9\}$, die erste Ziffer darf nicht 0 sein.
- Gehört das Fahrzeug allerdings einer Behörde an, so ist das Buchstabenfeld leer und der Ziffernblock 1-5 Ziffern lang. Dann darf der Ziffernblock ebenfalls nicht mit einer 0 anfangen und falls er 3 oder mehr Ziffern lang ist, auch nicht mit einer 4.
- Anmerkung: Es gibt weitere Sonderregelungen, welche in dieser Aufgabe ignoriert werden.

Geben Sie einen regulären Ausdruck für die gültigen Münchner Autokennzeichen an.

- c) Betrachten Sie folgenden NFA A :



Geben Sie einen regulären Ausdruck an, der die von A erkannte Sprache $L(A)$ erzeugt.

Hinweis: Sie müssen nicht das Verfahren aus der Vorlesung befolgen. Ein Vorgehen ist es, die vom NFA erkannte Sprache zunächst in Wörtern zu beschreiben und sich im Anschluss daran einen regulären Ausdruck dafür zu überlegen.

FSK4-2 Reguläre Ausdrücke: Geldscheine wechseln (0 Punkte)

Ein Geldscheinwechselautomat nimmt 10- und 20-Euro-Scheine als Eingabe und wechselt diese in 50-, 100- und 200-Euro-Scheine. Er erhält als Eingabe ein Wort über $\Sigma = \{10, 20\}$ und akzeptiert genau dann, wenn die Summe des Worts größer 0 und durch 50 teilbar ist. Zum Beispiel akzeptiert er 10 10 20 10 20 20 10 (da die Summe 100 ist), aber er akzeptiert 10 10 20 20 nicht (da die Summe 60 ist).

Geben Sie einen regulären Ausdruck an, der genau die durch den Geldscheinwechselautomaten akzeptierte Sprache erzeugt.

Hinweis: Es kann sinnvoll sein, zuerst einen endlichen Automaten zu konstruieren und dann den regulären Ausdruck zu entwerfen.

FSK4-3 Grammatik über Automaten zu Grammatik (2 Punkte)

Gegeben sei die Grammatik

$$G = (\{S, A, B, C\}, \{a, b\}, \{S \rightarrow aA \mid bB, A \rightarrow bB, B \rightarrow bC, C \rightarrow aC \mid a\}, S)$$

- Erzeugen Sie gemäß der Konstruktion aus der Vorlesung aus G einen NFA A mit $L(G) = L(A)$. Zeichnen Sie den Zustandsgraph von A .
- Erzeugen Sie mit der Potenzmengenkonstruktion aus A einen DFA B mit $L(B) = L(A)$. Zeichnen Sie den vom Startzustand erreichbaren Teil des Zustandsgraphen von B .
- Erzeugen Sie gemäß der Konstruktion aus der Vorlesung aus B eine Grammatik H mit $L(B) = L(H)$.
- Vergleichen Sie die Grammatiken G und H . Beschreiben Sie die Gemeinsamkeiten dieser Grammatiken, sowie ihre Unterschiede.

Überlegen Sie sich, wodurch diese Effekte zustande kommen.

FSK4-4 DNA-Analyse mit NFA (0 Punkte)

Diese Aufgabe handelt von der Analyse von Desoxyribonukleinsäure (DNS/DNA) mithilfe von NFA. DNA ist eine Abfolge der Basen Adenin, Thymin, Guanin und Cytosin, typischerweise mit A, T, G und C abgekürzt. Dementsprechend ist das Alphabet aller Automaten in dieser Aufgabe $\Sigma = \{A, C, G, T\}$.

- a) Um das Vorkommen einer Basensequenz zu finden, wird aus dieser Sequenz ein NFA erzeugt, der alle Wörter akzeptiert, in denen diese Sequenz als Teilwort vorkommt.

Zeichnen Sie den Zustandsgraph eines NFA B , der genau diejenigen Wörter akzeptiert, in denen $ACTC$ als Teilwort vorkommt.

Hinweis: Sie können (müssen aber nicht) dazu den regulären Ausdruck $(A|C|G|T)^*ACTC(A|C|G|T)^*$ verwenden, der genau diese Sprache akzeptiert.

Hinweis: Sie können auch einen DFA angeben, aber ein NFA ist übersichtlicher.

- b) Beim Kopieren von DNA kann es vorkommen, dass Fehler auftreten. Zum Beispiel kann eine Base durch eine andere ersetzt werden; es kann eine Base ausgelassen werden; es kann eine zusätzliche Base eingefügt werden; und es können auch komplexere Fehler auftreten. Zur Vereinfachung behandeln wir hier nur den Fall, dass eine Base durch eine andere ersetzt wird.

Aus einem NFA $D = (Z, \Sigma, \delta, S, E)$ kann ein NFA $F = (Z', \Sigma, \delta', S', E')$ erzeugt werden, der alle Wörter akzeptiert, die durch höchstens k fehlerhafte Ersetzungen aus D entsteht.

Dabei sind

- $Z' = Z \times \{0, \dots, k\}$
- $\delta'((q, i), a) = \{(q', i) \mid q' \in \delta(q, a)\} \cup \{(q', i+1) \mid (\exists b \in \Sigma. q' \in \delta(q, b) \wedge i+1 \leq k)\}$
- $S' = S \times \{0\} = \{(s, 0) \mid s \in S\}$
- $E' = E \times \{0, \dots, k\}$

Berechnen Sie mit der obigen Konstruktion einen NFA H aus B , der Wörter mit bis zu 2 Fehlern akzeptiert. Zeichnen Sie den Zustandsgraphen von H .

- c) Geben Sie an und begründen Sie, welche der folgenden Wörter von H akzeptiert werden. Prüfen Sie, ob Ihr Ergebnis korrekt ist, also ob die erkannten Wörter tatsächlich diejenigen sind, bei denen bis auf höchstens 2 Fehler das Wort $ACTC$ als Teilwort vorkommt.

$AAAACCCAAA, GAGGCGT, TAGCA, TCTCA$

- d) Begründen Sie, dass die Konstruktion aus b) korrekt ist, also tatsächlich für jeden NFA D und jedes k einen NFA F liefert, der maximal k Fehler zulässt.

Hinweis: Sie können auch als Vorüberlegung dies erst für den NFA H zeigen.