Klausur zur Vorlesung Automatentheorie und Formale Sprachen

Aufgabe 1 (10 Punkte)

Untersuchen Sie für jede der folgenden Sprachen, ob sie regulär und ob sie kontextfrei ist.

(a) Die Sprache aller $w \in \{a, b\}^*$, so daß für jeden Präfix v von w gilt: Die Anzahl der a's und die Anzahl der b's in v unterscheiden sich höchstens um drei.

(b) $L(G)$ mit Grammatik $G = (\{S\}, \{a, b\}, P, S)$ und $P = \{\, S \to aSb \mid SS \mid a \mid b \mid \varepsilon \,\}$

(c) $\{\, a^n b^m c^n \mid n, m \in \mathbf{N} \,\}$

(d) $\{\, a^{n^2} \mid n \in \mathbf{N} \,\}$

Lösung

(a) Regulär und somit kontextfrei durch Angabe eines DFA: Die Differenz zwischen der Anzahl der bisher gelesenen a's und der Anzahl der bisher gelesenen b's läßt sich einfach mit sieben Zuständen, die die Differenzen von -3 bis $+3$ repräsentieren, verwalten. Ein achter Zustand dient als Fangzustand; er stellt den einzigen nicht-akzeptierenden Zustand dar.

(b) Diese Sprache muß kontextfrei sein, weil sie die Sprache einer kontextfreien Grammatik ist. Vor allem ist sie aber auch regulär, weil sie nämlich $\{a, b\}^*$ ist: Alleine mit den Regeln $S \to SS \mid a \mid b \mid \varepsilon$ läßt sich schon jedes Wort aus $\{a, b\}^*$ erzeugen.

(c) Kontextfrei: Offenbar erzeugt die kontextfreie Grammatik mit den Produktionen $S \to aSc \mid B$ und $B \to bB \mid \varepsilon$ genau diese Sprache. Nicht regulär: Pumping-Lemma und Myhill-Nerode lassen sich hier beide recht gut anwenden, es geht aber geschickter. Der Homomorphismus h mit $h(a) = a$, $h(b) = \varepsilon$ und $h(c) = b$ bildet die Sprache auf die als nicht-regulär bekannte Sprache $\{\, a^n b^n \mid n \in \mathbf{N} \,\}$ ab.

(d) Diese Sprache ist nicht kontextfrei, wie sich mit dem Pumping-Lemma schnell zeigen läßt. Sei also n die Konstante aus dem Pumping-Lemma, $w = a^{n^2}$ und eine beliebige Zerlegung $w = a^i a^j a^k a^l a^m$ mit $j + k + l \leq n$ und $j + l \geq 1$ gegeben. Offenbar hat $a^i (a^j)^2 a^k (a^l)^2 a^m$ die Länge $n^2 + j + l \leq n^2 + n < n^2 + 2n + 1 = (n + 1)^2$ und ist somit nicht mehr in der Sprache enthalten. Also ist diese Sprache nicht kontextfrei.

Aufgabe 2 (10 Punkte)

Gegeben sei die Grammatik G mit $S \to aA \mid Bb \mid \varepsilon$, $A \to Sb$ und $B \to aS$.

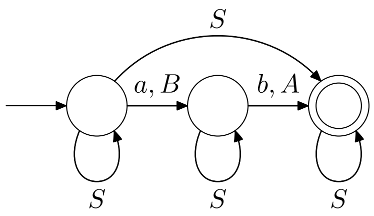
(a) Konstruieren Sie je einen NFA für $pre^*(\{ab\})$ und $pre^*(\{abab\})$. Welche der Aussagen $ab \in L(G)$ und $abab \in L(G)$ sind korrekt?

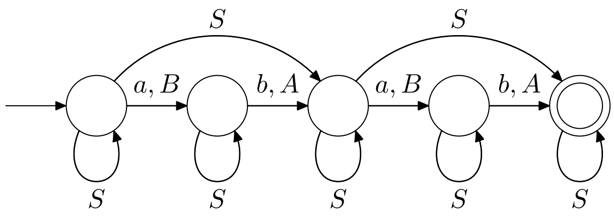
(b) Verwenden Sie die pre^*-Methode, um der folgenden Frage nachzugehen: Läßt sich aus S eine Satzform ableiten, die das Unterwort AA enthält?

(c) Ist G eine eindeutige Grammatik?

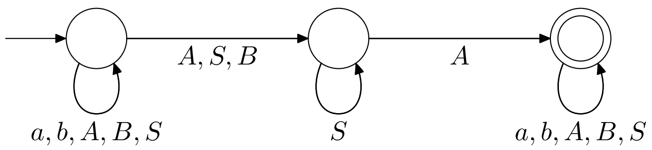
Lösung

(a) Die Sättigung des NFA für $\{ab\}$ zeigt schon nach wenigen Schritten, daß S vom Start- in den Endzustand führt. Es ist also $ab \in L(G)$. Für $abab$ hingegen ergibt sich dies nicht.





(b) Am gesättigten Automaten für $(a, b, A, B, S)^* AA (a, b, A, B, S)^*$ sieht man sofort, daß dies nicht möglich ist:



(c) Nein, denn ab läßt sich auf mehrere Weisen durch eine Linksableitung erzeugen, nämlich durch $S \Rightarrow aA \Rightarrow aSb \Rightarrow ab$ und $S \Rightarrow Bb \Rightarrow aSb \Rightarrow ab$.

Aufgabe 3 (10 Punkte)
Bringen Sie die folgende Grammatik erst in Chomsky-, dann in Greibach-Normalform:

$$S \to ABC \mid a, \quad A \to Sa, \quad B \to Sb, \quad C \to cS$$

Lösung
Zuerst bringen wir die Grammatik in Chomsky-Normalform und erhalten

$$S \to AX \mid a, \quad X \to BC, \quad A \to SR_a, \quad B \to SR_b, \quad C \to R_cS$$

mit $R_a \to a, R_b \to b, R_c \to c$, die wir ab jetzt verschweigen. Nun kümmern wir uns um S und erhalten zunächst

$$S \to SR_aX \mid a, \quad X \to BC, \quad A \to SR_a, \quad B \to SR_b, \quad C \to cS$$

und dann

$$S \to aZ \mid a, \quad Z \to aX \mid aXZ, \quad X \to BC, \quad A \to SR_a, \quad B \to SR_b, \quad C \to cS.$$

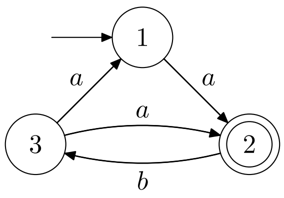
Abschließend betrachten wir A, B und X. Für A und B ist die Auflösung einfach, da die rechten Seiten mit S beginnen, S aber schon konvertiert ist. Danach kann auch X umgewandelt werden, und insgesamt ergibt sich

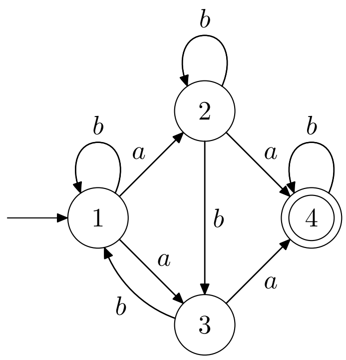
$$\begin{gathered} S \to aZ \mid a, \quad Z \to aX \mid aXZ, \quad C \to cS, \\ A \to aZR_a \mid aR_a, \quad B \to aZR_b \mid aR_b, \quad X \to aZR_bC \mid aR_bC. \end{gathered}$$

Aufgabe 4 (10 Punkte)

(a) Ermitteln Sie anhand einer geeigneten Methode einen regulären Ausdruck, der die von dem unten links (↙) abgebildeten Automaten erkannte Sprache beschreibt.

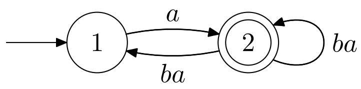
(b) Wandeln Sie den unten rechts (↘) abgebildeten NFA in einen DFA um.





Lösung

(a) Zustandselimination klappt hier natürlich ganz hervorragend. Sie liefert den unten abgebildeten Automaten und durch diesen den Ausdruck $a((ba)^+a)^*$ (oder auch $(a(ba)^+)^*a(ba)^*$).



Verwenden wir hingegen die $R^k_{i,j}$-Methode, so liefert $R^3_{1,2}$ den gesuchten Ausdruck. Es gilt $R^3_{1,2} = R^2_{1,2} + R^2_{1,3}(R^2_{3,3})^* R^2_{3,2}$. Für diese Komponenten erhalten wir durch Ablesen die Werte $R^2_{1,2} = a$, $R^2_{1,3} = ab$, $R^2_{3,3} = (aab + ab)^*$ und $R^2_{3,2} = a + aa$ und somit das Ergebnis $a + ab(aab + ab)^*(a + aa)$.

(b) Die Potenzmengenkonstruktion liefert einen feinen DFA: