Lösungsvorschlag zur Vorlesung Formale Sprachen, Automaten und Prozesse

Hinweis: In der Exkursionswoche finden keine Tutorien statt. Die Hausaufgaben zu diesem Blatt müssen daher erst in zwei Wochen abgegeben werden.

Aufgabe T15

Wir verwenden die (induktive) Definition von Unterscheidbarkeit, für einen Induktionsbeweis:

- Induktionsanfang: Seien q_1 und q_2 unterscheidbar, weil $q_1 \in F$, aber $q_2 \notin F$ ist. Dann gilt natürlich $\hat{\delta}(q_1, \varepsilon) \in F$ und $\hat{\delta}(q_1, \varepsilon) \notin F$. Analog verhält es sich, wenn $q_1 \notin F$ und $q_2 \in F$ sind.

- Induktionsschritt: Seien q_1 und q_2 unterscheidbar, weil ein $a \in \Sigma$ existiert, so daß $q_1' := \delta(q_1, a)$ und $q_2' := \delta(q_2, a)$ unterscheidbar sind. Nach Induktionsvoraussetzung existiert daher ein $w \in \Sigma^*$ für das Zustandspaar q_1' und q_2', so daß $\hat{\delta}(q_1', w) \in F$ und $\hat{\delta}(q_2', w) \notin F$ oder $\hat{\delta}(q_1', w) \notin F$ und $\hat{\delta}(q_2', w) \in F$. Wegen $\hat{\delta}(q_1, aw) = \hat{\delta}(\delta(q_1, a), w) = \hat{\delta}(q_1', w)$ und $\hat{\delta}(q_2, aw) = \hat{\delta}(\delta(q_2, a), w) = \hat{\delta}(q_2', w)$ folgt sofort die Aussage.

Aufgabe T16

a) Beweis durch Widerspruch. Sei die Sprache der korrekten bedingten Ausdrücke regulär. Dann gibt es nach dem Pumping-Lemma eine Zahl n mit den beschriebenen Eigenschaften. Sei $w = \text{if}^n\text{A(then A fi)}^n$ und $xyz = w$ eine beliebige Zerlegung von w mit $|xy| \leq n$ und $|y| > 0$. Somit ist xy ein Präfix von if^n und xy^2z kann kein korrekter bedingter Ausdruck sein, da er unterschiedliche Anzahlen von if und fi enthält.

b) Analog zu oben, sei $w = (^n[+1)]^n$. Somit gilt für jede Zerlegung $xyz = w$ mit $|xy| \leq n$ und $|y| > 0$, daß xy ein Präfix von $(^n$ ist. Somit ist xy^2z kein korrekt geklammerter Ausdruck, da sich die Zahl der öffnenden und schließenden Klammern unterscheiden.

c) Angenommen, die Sprache L_c dieser Aufgabe sei regulär. Nach Pumping-Lemma existiert dann ein $n \in \mathbf{N}$, sodaß für alle $w \in L_c$ mit $|w| \geq n$ eine Zerlegung $w = xyz$ existiert, mit $|xy| \leq n$, $|y| \geq 1$ und $xy^kz \in L_c$ für alle $k \in \mathbf{N}$.

Sei $w = 1^n - 1^n + 1$. Wertet man diesen Ausdruck aus, so ergibt sich offensichtlich 1, also ist w in der Sprache enthalten. Außerdem ist offensichtlich $|w| > n$.

Sei $xyz = w$ eine beliebige Zerlegung von w mit $|xy| \leq n$ und $|y| \geq 1$. Dann ist $xy^0z = xz = 1^{n-|y|} - 1^n + 1$. Offensichtlich ist aber xz nicht in L_c, da der zugehörige Wert kleiner als 1 ist. Dies ist ein Widerspruch zur Aussage des Pumping-Lemmas, also kann L_c nicht regulär sein.

Aufgabe T17

Wir wandeln die regulären Ausdrücke mit der Tompson-Konstruktion in ϵ-NFAs $M_1, \dots M_k$ um. Wir erweitern das Eingabealphabet um neue Symbole $r_1, \dots, r_n$. Wir vereinigen die ϵ-NFAs zu einem neuen ϵ-NFA M über dem größeren Alphabet wie folgt:

- Wir fügen einen neuen Startzustand ein, von dem man per ϵ-Transition zu jedem Startzustand von $M_1, \dots, M_k$ kommen kann.
- Wir fügen einen neuen Endzustand q_F ein. Von jedem Endzustand in M_l fügen wir eine neue r_l-Transition zu q_F ein. Alle Endzustände von $M_1, \dots, M_k$ werden daraufhin zu normalen Zuständen.

Nun wandeln wir den ϵ-NFA in einen minimalen DFA M^* um. Dies geschieht, indem wir einen äquivalenten NFA erstellen, die Potenzmengenkonstruktion ausführen und den resultierenden Automaten minimieren. Sei δ die Transitionsfunktion von M^*. Nun ist $\hat{\delta}(q_0, u)$ ein Zustand von dem eine Transition r_l möglich ist, genau dann wenn der reguläre Ausdruck r_l den Prefix u erkennt.

Sei $w = w_1$ das Eingabewort. Wir erzeugen ein leeres Array A der Länge $|w_1|$ und simulieren M^* für $|w_1|$ Schritte auf dem Wort. Wenn M^* im $j$ten Schritt in einen Zustand kommt, von dem eine Transition $r \in R$ möglich ist, schreiben wir r an die $j$te Stelle im Array. Nachdem wir den Automaten komplett simuliert haben, durchlaufen wir A von hinten, bis wir den ersten nichtleeren Eintrag finden.

Falls es keinen solchen Eintrag gibt, gibt es keinen regulären Audruck $r \in R$, der einen Prefix von w_1 erkennt. Wir terminieren.

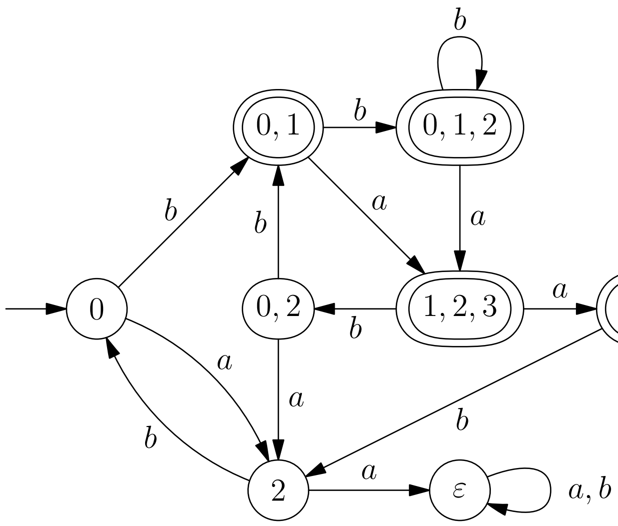
Andernfalls können wir annehmen der erste nichtleere Eintrag von hinten lautet r und ist an der $j$ten Stelle des Arrays. Wir definieren $q_1 = r$ und u_1 als den Prefix der Länge j von w_1. Wir geben q_1 und u_1 aus.

Wir erhalten das nächste Wort w_2, indem wir den Prefix u_1 von w_1 entfernen. Wir erzeugen wieder ein leeres Array der Länge $|w_2|$ und simulieren M^* auf dem Wort w_2. Dieses Verfahren setzen wir fort, und erhalten so $q_1, \dots q_k$ und $u_1, \dots u_k$. Der Algorithmus terminiert, sobald ein Wort $w_{k+1} = \epsilon$ erreicht wird. In jedem Durchlauf wird er längste Prefix des aktuellen Wortes gefunden, der von einem regulären Audruck erkann wird. Das Verfahren ist somit korrekt.

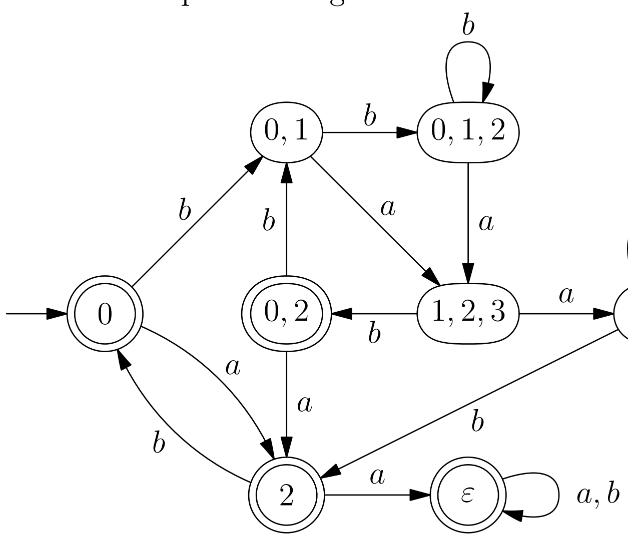
Aufgabe H10 (2+18 Punkte)

a) Ja, durch den Pfad $(0 \xrightarrow{b} 0 \xrightarrow{b} 0 \xrightarrow{b} 0 \xrightarrow{a} 2 \xrightarrow{b} 0 \xrightarrow{b} 1 \xrightarrow{a} 1 \xrightarrow{a} 1 \xrightarrow{a} 1)$

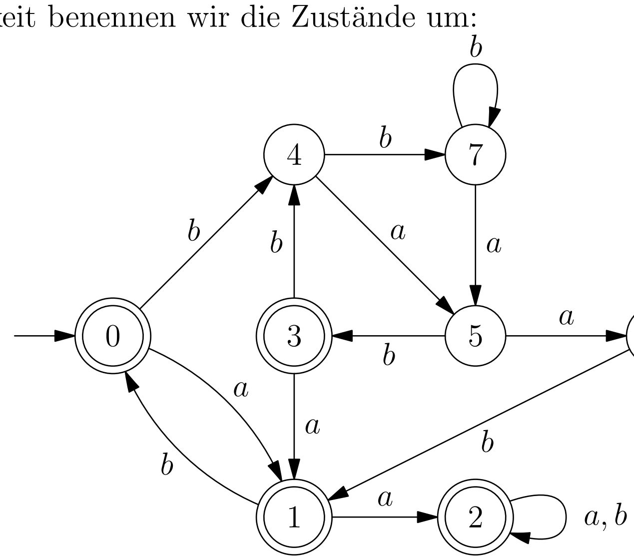
b) Durch die Potenzmengenkonstruktion kann der zu dem NFA äquivalente DFA gewonnen werden:



Nun kann die Negation der Sprache durch Vertauschen der Endzustände erreicht werden. Der resultierende Automat akzeptiert nun genau die unsicheren Öffnungssequenzen.



Zur Übersichtlichkeit benennen wir die Zustände um:



Um den regulären Ausdruck nicht zu groß werden zu lassen, führen wir die Minimierung durch den Markierungsalgorithmus aus der Vorlesung durch:

	7	6	5	4	3	2	1
0	X	X	X	X		X	X
1	X	X	X	X	X	X	
2	X	X	X	X	X		
3	X	X	X	X			
4		X	X				
5	X	X					
6	X						

$0,3 \xrightarrow{b} 4,4$
$\quad\;\; \xrightarrow{a} 1,1$

$0,2 \xrightarrow{b} 2,4$

$0,1 \xrightarrow{b} 0,4$

$1,3 \xrightarrow{b} 2,4$
$\quad\;\; \xrightarrow{a} 1,2$

$1,2 \xrightarrow{b} 0,2$

$2,3 \xrightarrow{b} 2,4$

$4,7 \xrightarrow{b} 7,7$
$\quad\;\; \xrightarrow{a} 5,5$

$4,6 \xrightarrow{b} 1,7$

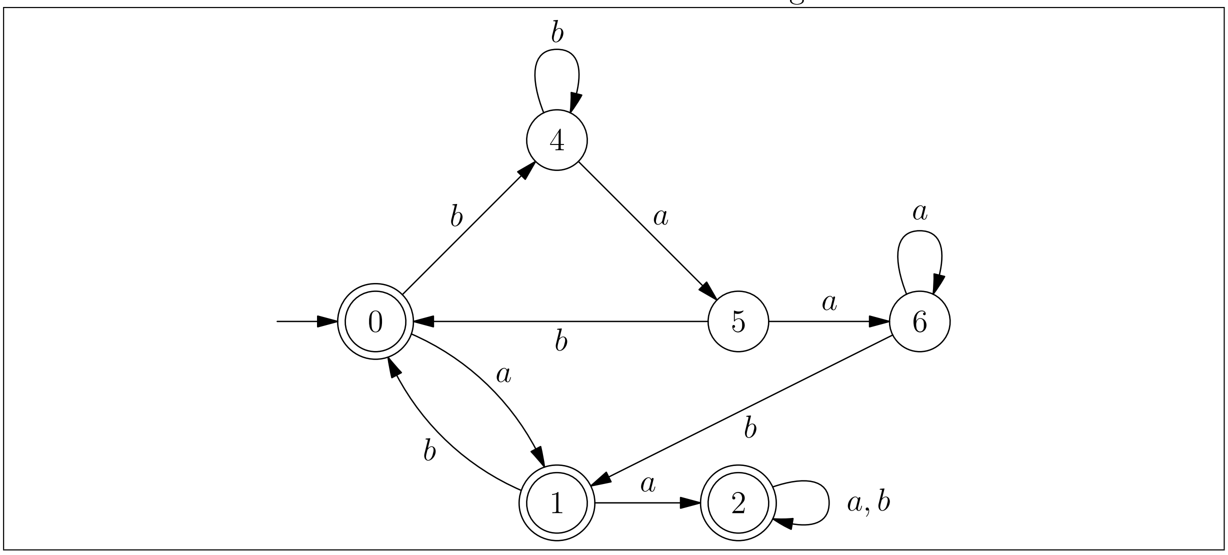
$4,3 \xrightarrow{b} 3,7$

$5,7 \xrightarrow{b} 3,7$

$5,7 \xrightarrow{b} 1,3$

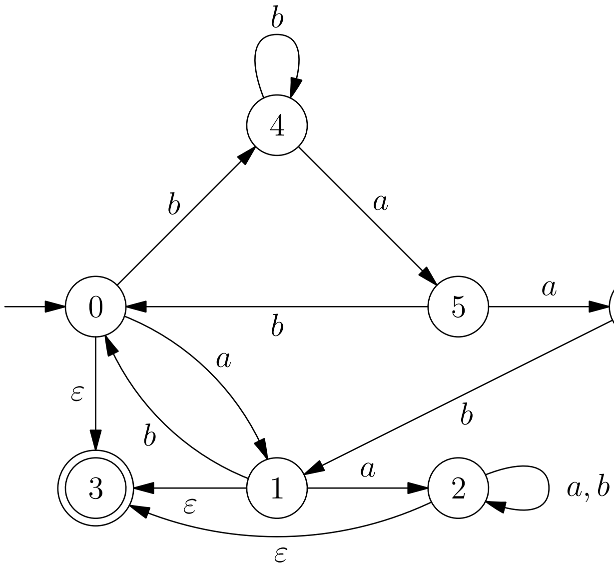
$6,7 \xrightarrow{b} 1,7$

Somit können die Zustände 4 und 7 sowie 0 und 3 vereinigt werden:

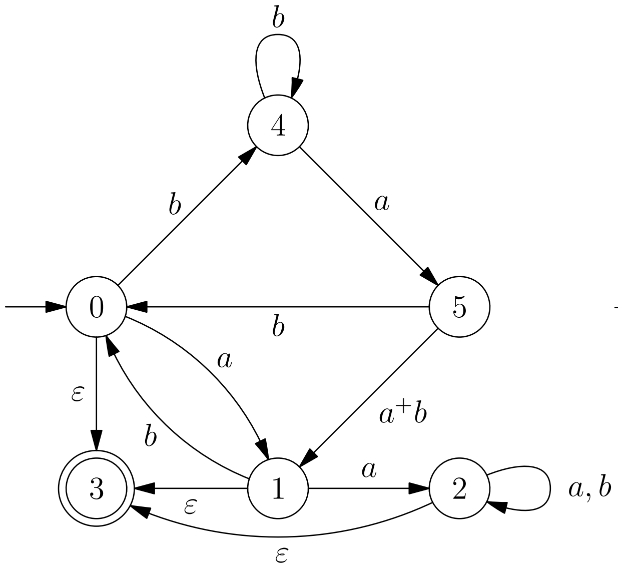


Dies ist der DFA, der das Komplement der Sprache von dem NFA aus der Aufgabenstellung erkennt.

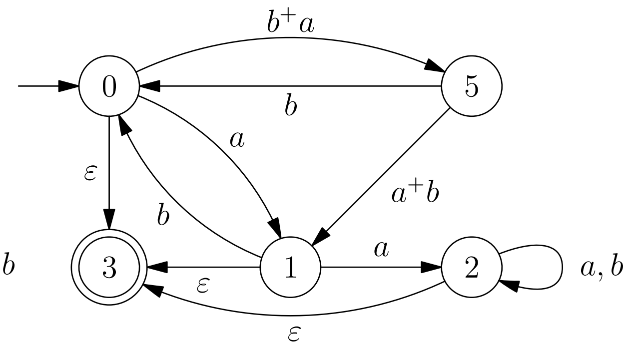
Um den regulären Ausdruck konstruieren zu können, führen wir einen eindeutigen Endzustand ein, der von den ursprünglichen Endzuständen über ε-Kanten erreicht wird:



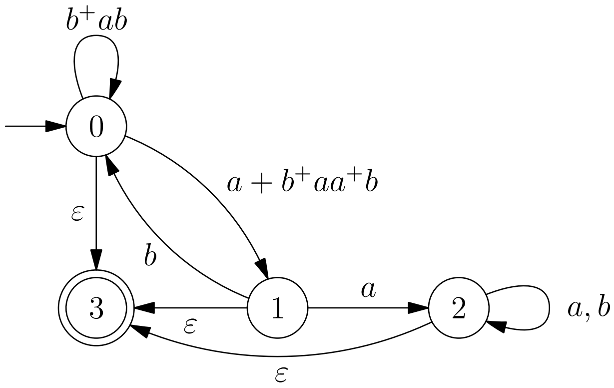
Nun können wir schrittweise Knoten aus dem Graphen entfernen und dabei die betroffenen Kanten mit den entsprechenden regulären Ausdrücken markieren.



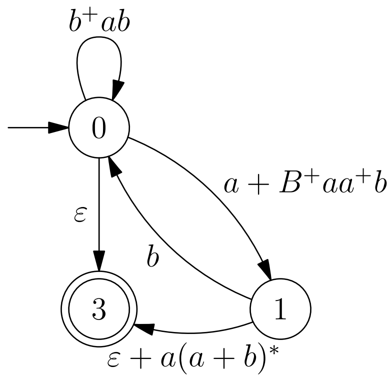
Zustand 6 entfernt.



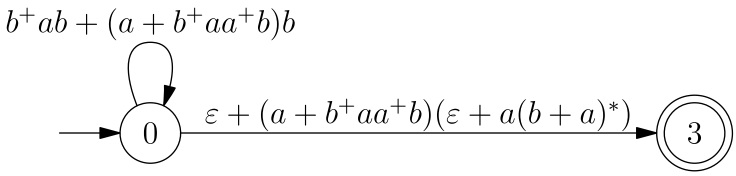
Zustand 4 entfernt.



Zustand 5 entfernt.



Zustand 2 entfernt.



Zustand 1 entfernt.

$$\boxed{(b^+ab + (a + b^+aa^+b)b)^*(\varepsilon + (a + b^+aa^+b)(\varepsilon + a(b + a)^*))}$$

Ein regulärer Ausdruck, der die falsch kodierten Sequenzen beschreibt.

Aufgabe H11 (10 Punkte)

Angenommen L sei regulär, nach dem Pumping-Lemma existiert dann ein $n \in \mathbf{N}$, so daß für alle $w \in L$ mit $|w| \geq n$ eine Zerlegung $w = xyz$ existiert, mit $|xy| \leq n$, $|y| \geq 1$ und $xy^kz \in L$ für alle $k \in \mathbf{N}$.

Wir definieren uns $u = l^9$ für eine 90 Grad Drehung. Sei $w = g^nugug^{(n+1)}ugg$. Wertet man diesen Ausdruck aus, so ergibt sich eine n Zentimeter lange Linie (nehmen wir ohne beschränkung der Allgemeinheit an, daß die erste Linie nach rechts führt), gefolgt von einer 90 Grad Drehung, eine ein Zentimeter langen Line nach unten, eine weitere 90 Grad drehung gefolgt von einer $n+1$ Zentimeter langen Linie nach links. Der Stift ist nun einen Zentimeter links von dem ursprünglichen Startpunkt entfernt, das heißt wenn man jetzt zwei Zentimeter nach oben geht, kreuzt sich die Linie nicht, also ist w in der Sprache enthalten. Außerdem ist offensichtlich $|w| > n$.

Sei $xyz = w$ eine beliebige Zerlegung von w mit $|xy| \leq n$ und $|y| \geq 1$. Dann ist besteht xy nur aus „bewege dich einen Zentimeter (g)" Befehlen. Insbesondere enthält y mindestens ein g. Wählen wir nun $i = 2$, dann gilt $xy^2 = g^m$ mit $m \geq 2$. Damit ist $w' = xy^2z = g^{m'}ugug^{(n+1)}ugg$ mit $m' \geq n+2$. Was aber zu einer sich kreuzenden Linie führt, da man weiter nach Rechts geht als man wieder nach Links zurück geht. Also folgt $w' \notin L$. Dies ist ein Widerspruch zur Aussage des Pumping-Lemmas, also kann L nicht regulär sein.