technische universität dortmund

Thomas Schwentick

Gaetano Geck, Lutz Oettershagen,
Christopher Spinrath, Marco Wilhelm

Lehrstuhl 1
Logik in der Informatik

SoSe 2018 | **Übungsblatt 2** | **19.04.2018**

Aufgabe 2.1 [Automatenkonstruktion und -interpretation] — 5 Punkte

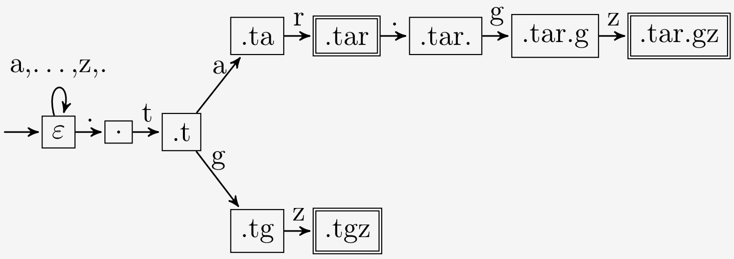
a) Hannah arbeitet seit einiger Zeit an einem Webportal. Dieses soll nun so erweitert werden, dass Lehrmaterialien in Form komprimierter oder unkomprimierter Archivdateien hochgeladen werden können. Erlaubt sind TAR-Archive, eventuell mit GZIP komprimiert. Gültig sind demnach die Dateiendungen `.tar`, `.tar.gz`, `.tgz`. Der Einfachheit halber sei vorausgesetzt, dass alle Dateinamen ausschließlich Kleinbuchstaben und Punkte beinhalten.

Konstruieren Sie einen ε-NFA über dem Alphabet $\{a, \dots, z, .\}$ mit möglichst wenigen Zuständen, der genau die Dateinamen mit gültigen Dateiendungen akzeptiert. **(2,5 Punkte)**

b) Konstruieren Sie einen DFA mit maximal sieben Zuständen, der genau die Wörter über dem Alphabet $\{a, b\}$ akzeptiert, die weder mit aa noch mit bb enden. **(2,5 Punkte)**

Lösung:

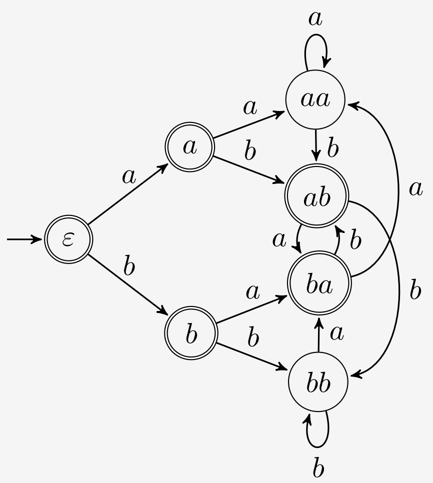
a) Ein NFA, der genau die Dateinamen mit den gültigen Endungen akzeptiert ist nachfolgend angegeben. Jeder Zustand entspricht einem längesten Suffix des bisher gelesenen Wortes, das zu einer gültigen Endung führen kann.



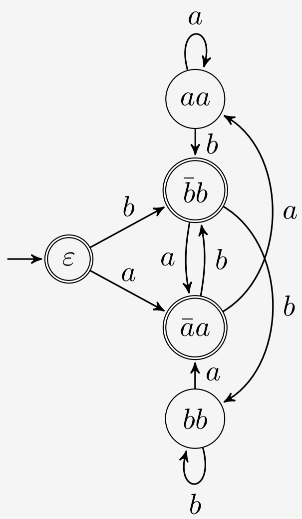
Ein äquivalenter NFA mit weniger Zuständen (ohne `.tg` und `.tgz`) hat eine Transition (`.t`, g, `.tar.g`).

b) Der folgende DFA akzeptiert genau die Wörter über $\{a, b\}$, die weder mit aa noch mit bb enden. Die Zustände repräsentieren die zwei zuletzt gelesenen Zeichen (aa, ab, ba, bb), sofern schon vorhanden, oder das einzige bisher gelesene Zeichen (a oder b) oder den Umstand, das bisher noch kein Zeichen gelesen wurde (ε).

Die Transitionen entsprechen der Veränderung der zu merkenden Zeichen: Wurde beispielsweise bisher das Wort $abba$ gelesen, dann befindet sich der Automat in Zustand ba und geht nach Lesen eines weiteren Zeichens a, also für das Gesamtwort $abbaa$, in den Zustand aa über.



Der Automat kann vereinfacht werden, indem die Zustände a und ba bzw. die Zustände b und ab miteinander identifiziert werden (vgl. die Zustände $\bar{a}a$ und $\bar{b}b$). Dem liegt die Idee zugrunde, dass vor dem letzten gelesenen Zeichen (jeweils a oder b) nicht bereits dasselbe Zeichen gelesen wurde (also entweder gar kein Zeichen oder das andere).

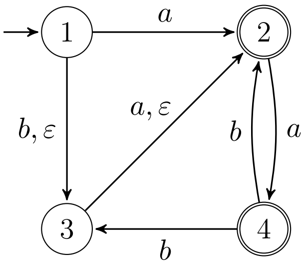


Aufgabe 2.2 [Äquivalenz der Modelle] 8 Punkte

a) RE ⇝ ε-NFA: Konstruieren Sie einen zum regulären Ausdruck $\alpha = ((a + \varepsilon)bb)^*$ äquivalenten ε-NFA.

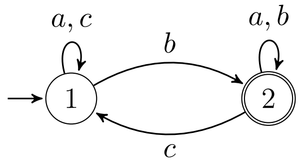
Gehen Sie dabei genau nach dem in der Beweisskizze zu Proposition 2.2 vorgestellten induktiven Prinzip („Baukastenprinzip“) vor. Fügen Sie insbesondere, alle ε-Transitionen ein. Markieren Sie außerdem, welche Komponenten des ε-NFAs welchen Teilausdrücken von α entsprechen. **(2 Punkte)**

b) ε-NFA ⇝ DFA: Konstruieren Sie den Potenzmengen-Automaten zu dem folgenden ε-NFA über dem Alphabet $\{a, b\}$. Beschränken Sie sich auf die vom Startzustand aus erreichbaren Zustände.



Begründen Sie Ihre Wahl des Startzustandes und Ihre Wahl der akzeptierenden Zustände. Geben Sie insbesondere die Menge ε-closure(1) an. **(3 Punkte)**

c) **DFA $\rightsquigarrow$ RE:** Konstruieren Sie einen zu folgendem DFA $\mathcal{A}$ äquivalenten regulären Ausdruck. Gehen Sie dabei nach dem in der Beweisskizze zu Proposition 3.3 vorgestellten Verfahren vor.



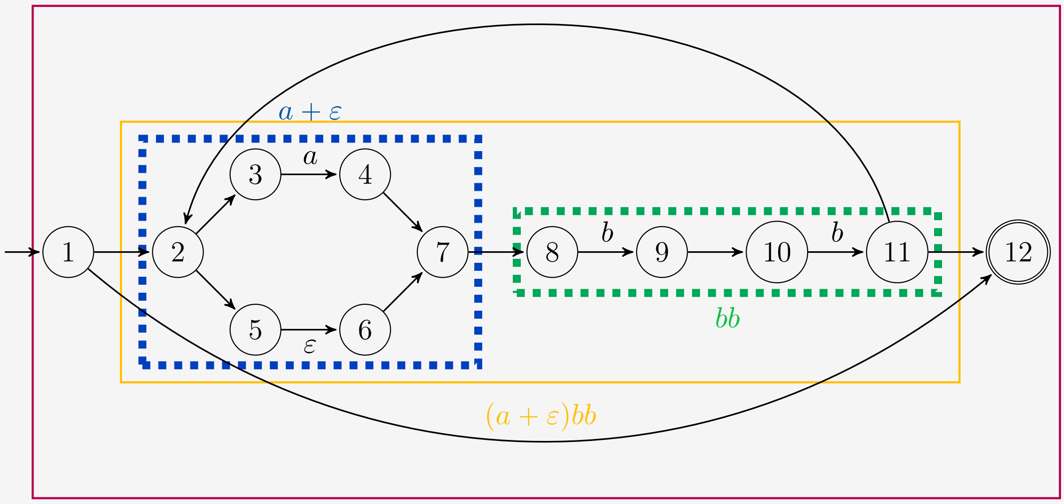
Hinweis

Nach Bestimmung eines Ausdrucks $\alpha_{i,j}^{k}$ dürfen Sie diesen äquivalent vereinfachen und mit diesem vereinfachten Ausdruck die Berechnung weiterführen. Außerdem müssen Sie einen Ausdruck $\alpha_{i,j}^{k}$ nur bestimmen, wenn Sie ihn für die Konstruktion des Ausdrucks, der äquivalent zu $\mathcal{A}$ ist, oder zur Bestimmung eines anderen Ausdrucks $\alpha_{i',j'}^{k'}$ benötigen.

(3 Punkte)

Lösung:

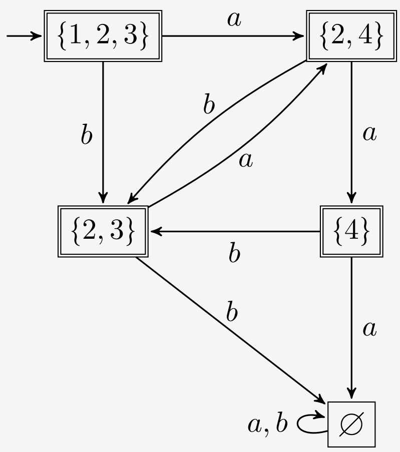
a) Der ε-NFA, der nach dem induktiven Prinzip aus dem Ausdruck $\big((a+\varepsilon)bb\big)^*$ entsteht, ist nachfolgend abgebildet. Die unbeschrifteten Kanten entsprechen ε-Transitionen.



$\big((a + \varepsilon)bb\big)^*$

b) Zunächst bestimmen wir den ε-Abschluss der entsprechenden Zustandsmengen. Der Startzustand des Potenzmengenautomaten ist ε-closure$(1) = \{1, 2, 3\}$, da diese drei Zustände im Ausgangsautomaten vom dortigen Startzustand 1 ohne Lesen eines Zeichens erreicht werden können.

Insgesamt ergibt sich der folgende Potenzmengenautomat



Akzeptierend sind genau die Zustände, die mindestens einen akzeptierenden Zustand (2 bzw. 4) des Ausgangsautomaten enthalten.

c) Jeder Ausdruck $\alpha^0_{i,j}$ beschreibt genau die Wörter, die von Zustand i in Zustand j führen *ohne einen Zwischenzustand zu nutzen.*

$$\alpha^0_{1,1} = \varepsilon + a + c$$
$$\alpha^0_{1,2} = b$$
$$\alpha^0_{2,1} = c$$
$$\alpha^0_{2,2} = \varepsilon + a + b$$

Jeder Ausdruck $\alpha^1_{i,j}$ beschreibt genau die Wörter, die von Zustand i in Zustand j führen, wobei ausschließlich Zustand 1 als Zwischenzustand genutzt werden kann (nicht aber genutzt werden muss).

$$\begin{aligned}
\alpha^1_{1,1} &= (\varepsilon + a + c) + (\varepsilon + a + c)(\varepsilon + a + c)^*(\varepsilon + a + c) && \equiv (a + c)^* \\
\alpha^1_{1,2} &= b + (\varepsilon + a + c)(\varepsilon + a + c)^* b && \equiv (a + c)^* b \\
\alpha^1_{2,1} &= c + c(\varepsilon + a + c)^*(\varepsilon + a + c) && \equiv c(a + c)^* \\
\alpha^1_{2,2} &= (\varepsilon + a + b) + c(\varepsilon + a + c)^* b && \equiv \varepsilon + a + b + c(a + c)^* b
\end{aligned}$$

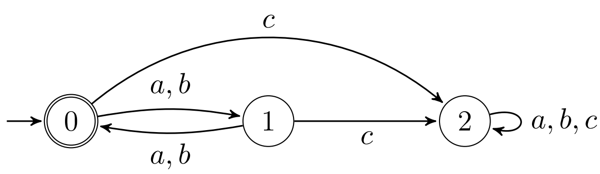
Zuletzt beschreibt der reguläre Ausdruck $\alpha^2_{1,2}$ genau die Wörter, die vom Startzustand 1 zum einzigen akzeptierenden Zustand 2 des Ausgangsautomaten führen und dabei über beliebige Zwischenzustände führen können.

$$\alpha^2_{1,2} = (a + c)^* b + (a + c)^* b(\varepsilon + a + b + c(a + c)^* b)^*(\varepsilon + a + b + c(a + c)^* b)$$

Da 2 der einzige akzeptierende Zustand des Automaten ist, ist der gesuchte reguläre Ausdruck $\alpha = \alpha^2_{1,2}$ (mit $L(\alpha) = L(\mathcal{A})$).

Aufgabe 2.3 [Sprachäquivalenzen] **2 Punkte**

Seien die Sprache $L = \{w \in \Sigma^* \mid \#_c(w) = 0 \text{ und } |w| \text{ hat gerade Länge}\}$ und der folgende Automat $\mathcal{A}$, jeweils über dem Alphabet $\Sigma = \{a, b, c\}$, gegeben.



Zeigen Sie $L(\mathcal{A}) = L$. Geben Sie dazu für jeden Zustand $q \in \{0, 1, 2\}$ eine Sprache $W_q \subseteq \Sigma^*$ an und beweisen Sie, dass W_q genau die Wörter, die vom Startzustand 0 in den Zustand q führen, enthält. Beispielsweise ist $W_2 = \{w \in \Sigma^* \mid \#_c(w) > 0\}$ die Sprache aller Wörter, die mindestens ein c enthalten.

Lösung:

Die Sprachen W_0 und W_1 enthalten jeweils genau die Wörter gerader bzw. ungerader Länge, die ausschließlich die Zeichen a und b enthalten. Die Mengen W_0, W_1 und W_2 beschreiben also eine Partition von Σ^*:

$$\begin{aligned} W_0 &= \{w \in \Sigma^* \mid \#_c(w) = 0 \text{ und } |w| \text{ ist gerade}\}, \\ W_1 &= \{w \in \Sigma^* \mid \#_c(w) = 0 \text{ und } |w| \text{ ist ungerade}\}, \\ W_2 &= \{w \in \Sigma^* \mid \#_c(w) > 0\}. \end{aligned}$$

Durch Induktion nach der Länge $|w|$ von w beweisen wir, dass für jedes $w \in \Sigma^*$ gilt:
Es gilt $w \in W_q$ für $q \in \{0, 1, 2\}$ *genau dann, wenn* $\delta^*(0, w) = q$ gilt.
Genau genommen, zeigen wir nur
Es gilt $w \in W_q$ für $q \in \{0, 1, 2\}$ dann, *wenn* $\delta^*(0, w) = q$ gilt.
Die Implikation in die Gegenrichtung folgt dann daraus, dass der Automat $\mathcal{A}$ deterministisch ist, es für jedes Wort $w \in \Sigma^*$ also genau einen Zustand q mit $\delta^*(0, w) = q$ gibt.

Induktionsanfang ($|w| = 0$): Es gilt $w = \varepsilon \in W_0$. Nach Definition der erweiterten Transitionsfunktion gilt trivialerweise $\delta^*(q, \varepsilon) = q$ für jeden Zustand q, insbesondere also $\delta^*(0, \varepsilon) = 0$.

Induktionsschritt ($|w| > 0$): Da w mindestens ein Zeichen enthält, lässt es sich schreiben als $w = u\sigma$ für ein Zeichen $\sigma \in \{a, b, c\}$ und ein Wort $u \in \{a, b, c\}^*$ mit $|u| < |w|$. Sei $p = \delta^*(0, u)$ der Zustand, in den der Automat $\mathcal{A}$ nach Lesen von u gelangt. Nach Induktionsvoraussetzung gilt $u \in W_p$. Wir unterscheiden drei Fälle, abhängig vom aktuellen Zustand p und vom nächsten Zeichen σ:

1. Fall ($\sigma = c$): Dann gilt $\#_c(w) = \#_c(u\sigma) > 0$, also $w \in W_2$. Außerdem lässt sich dem Transitionsdiagramm entnehmen, dass $\delta(q, c) = 2$ für jeden Zustand $q \in \{0, 1, 2\}$ gilt. Damit folgt $\delta^*(0, w) = \delta(\delta^*(u), \sigma) = \delta(p, \sigma) = 2$.

2. Fall ($\sigma \neq c$ und $p = 2$): Dann gilt $u \in W_2$, also $\#_c(u) > 0$. Ebenso gilt $\#_c(u\sigma) > 0$, also $w \in W_2$, da sich die Anzahl der c durch Lesen von σ nicht verringert. Dem Transitionsdiagramm lässt sich $\delta(2, \tau) = 2$ für jedes $\tau \in \Sigma$ entnehmen. Insbesondere gilt also $\delta(2, \sigma) = 2$ und somit $\delta^*(0, w) = \delta(p, \sigma) = 2$.

3. Fall ($\sigma \neq c$ und $p \neq 2$): Offenbar gilt dann $\sigma \in \{a, b\}$ und $p \in \{0, 1\}$. Wir betrachten den Fall $p = 0$ (für $p = 1$ kann analog argumentiert werden). Aus $u \in W_0$ folgt, dass u ausschließlich die Zeichen a und b enthält und gerade Länge hat. Entsprechend ist $w = u\sigma$ ungerader Länge und enthält ebenfalls auschließlich die Zeichen a und b, das heißt, es gilt $w \in W_1$. Dem Transitionsdiagramm lässt sich $\delta(0, \tau) = 1$ für jedes $\tau \in \{a, b\}$ entnehmen, insbesondere also $\delta(0, \sigma) = 1$. Somit gilt $\delta^*(0, w) = \delta(\delta^*(u), \sigma) = \delta(0, \sigma) = 1$.

In jedem Fall gilt also $w \in W_q$ dann, wenn $\delta^*(0, w) = q$ gilt. Da Zustand 0 der einzige akzeptierende Zustand von $\mathcal{A}$ ist, gilt $L(\mathcal{A}) = L = W_0$, wie behauptet.