SS 2006

Einführung in die Informatik IV

Ernst W. Mayr

Fakultät für Informatik
TU München

http://www14.in.tum.de/lehre/2006SS/info4/

Sommersemester 2006

3.3 Äquivalenz von NFA und DFA

Satz 30

Für jede von einem nichtdeterministischen endlichen Automaten akzeptierte Sprache L gibt es auch einen deterministischen endlichen Automaten M mit

$$L = L(M)\ .$$

Beweis:

Sei $N = (Q, \Sigma, \delta, S, F)$ ein NFA.

Definiere

1. $M' := (Q', \Sigma, \delta', q'_0, F')$
2. $Q' := \mathcal{P}(Q) \setminus \{\emptyset\}$ ($\mathcal{P}(Q)$ Potenzmenge von Q)
3. $\delta'(Q'', a) := \bigcup_{q' \in Q''} \delta(q', a)$ für alle $Q'' \in Q'$, $a \in \Sigma$
4. $q'_0 := S$
5. $F' := \{Q'' \subseteq Q;\ Q'' \cap F \neq \emptyset\}$

Also

NFA N:	Q	Σ	δ	S	F
DFA M':	2^Q	Σ	δ'	S	F'

Es gilt:

$$\begin{aligned} w \in L(N) &\Leftrightarrow \hat{\delta}(S, w) \cap F \neq \emptyset \\ &\Leftrightarrow \widehat{\delta'}(q'_0, w) \in F' \\ &\Leftrightarrow w \in L(M'). \end{aligned}$$

□

Der zugehörige Algorithmus zur Überführung eines NFA in einen DFA heißt Teilmengenkonstruktion, Potenzmengenkonstruktion oder Myhill-Konstruktion.

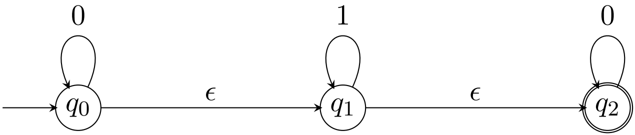
3.4 NFA's mit ϵ-Übergängen

Definition 31

Ein (nichtdeterministischer) endlicher Automat A mit ϵ-Übergängen ist ein 5-Tupel analog zur Definition des NFA mit

$$\delta : Q \times (\Sigma \uplus \{\epsilon\}) \to \mathcal{P}(Q) \setminus \{\emptyset\} \; .$$

Ein ϵ-Übergang wird ausgeführt, ohne dass ein Eingabezeichen gelesen wird. Wir setzen o.B.d.A. voraus, dass A nur einen Anfangszustand hat.



Definiere für alle $a \in \Sigma$

$$\bar{\delta}(q, a) := \hat{\delta}(q, \epsilon^* a \epsilon^*) \ .$$

Satz 32

$$w \in L(A) \Leftrightarrow \hat{\bar{\delta}}(S, w) \cap F \neq \emptyset \ .$$

Beweis:

Hausaufgabe! $\square$

3.5 Entfernen von ϵ-Übergängen

Satz 33

Zu jedem nichtdeterministischen endlichen Automaten A mit ϵ-Übergängen gibt es einen nichtdeterministischen endlichen Automaten A' ohne ϵ-Übergänge, so dass gilt:

$$L(A) = L(A')$$

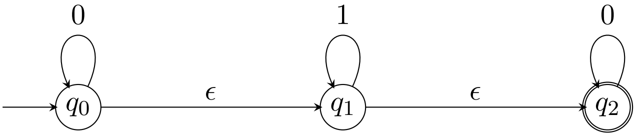
Beweis:

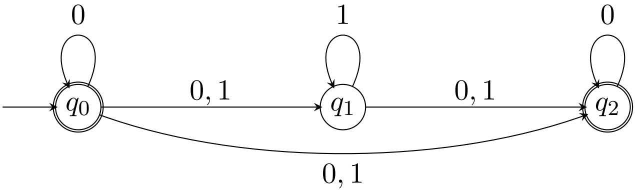
Ersetze δ durch $\bar{\delta}$ und F durch F' mit

$$F' = \begin{cases} F & \epsilon \notin L(A) \\ F \cup \{q_0\} & \epsilon \in L(A) \end{cases}$$

□

Beispiel 34





3.6 Endliche Automaten und reguläre Sprachen

Zusammenfassend ergibt sich:

Satz 35

Die Familie der regulären Sprachen (Chomsky-3-Sprachen) ist identisch mit der Familie der Sprachen, die

- *von DFA's akzeptiert/erkannt werden,*
- *von NFA's akzeptiert werden,*
- *von NFA's mit ϵ-Übergängen akzeptiert werden.*

Beweis:

Wie soeben gezeigt. $\square$

3.7 Reguläre Ausdrücke

Reguläre Ausdrücke sollen eine kompakte Notation für spezielle Sprachen sein, wobei endliche Ausdrücke hier auch unendliche Mengen beschreiben können.

Definition 36

Reguläre Ausdrücke sind induktiv definiert durch:

1. $\emptyset$ ist ein regulärer Ausdruck.
2. ϵ ist ein regulärer Ausdruck.
3. Für jedes $a \in \Sigma$ ist a ist ein regulärer Ausdruck.
4. Wenn α und β reguläre Ausdrücke sind, dann sind auch (α), $\alpha\beta$, $(\alpha|\beta)$ (hierfür wird oft auch $(\alpha + \beta)$ geschrieben) und $(\alpha)^*$ reguläre Ausdrücke.
5. Nichts sonst ist ein regulärer Ausdruck.

Zu einem regulären Ausdruck γ ist die zugehörige Sprache $L(\gamma)$ induktiv definiert durch:

Definition 37

1. Falls $\gamma = \emptyset$, so gilt $L(\gamma) = \emptyset$.
2. Falls $\gamma = \epsilon$, so gilt $L(\gamma) = \{\epsilon\}$.
3. Falls $\gamma = a$, so gilt $L(\gamma) = \{a\}$.
4. Falls $\gamma = (\alpha)$, so gilt $L(\gamma) = L(\alpha)$.
5. Falls $\gamma = \alpha\beta$, so gilt
$$L(\gamma) = L(\alpha)L(\beta) = \{uv;\ u \in L(\alpha), v \in L(\beta)\}\ .$$
6. Falls $\gamma = (\alpha \mid \beta)$, so gilt
$$L(\gamma) = L(\alpha) \cup L(\beta) = \{u;\ u \in L(\alpha) \vee u \in L(\beta)\}\ .$$
7. Falls $\gamma = (\alpha)^*$, so gilt
$$L(\gamma) = L(\alpha)^* = \{u_1 u_2 \ldots u_n;\ n \in \mathbb{N}_0, u_1, \ldots, u_n \in L(\alpha)\}\ .$$

Beispiel 38

Sei das zugrunde liegende Alphabet $\Sigma = \{0, 1\}$.

- alle Wörter, die gleich 0 sind oder mit 00 enden:

$$(0 \mid (0 \mid 1)^*00)$$

- alle Wörter, die 0110 enthalten:

$$(0|1)^*0110(0|1)^*$$

- alle Wörter, die eine gerade Anzahl von 1'en enthalten:

$$(0^*10^*1)^*0^*$$

- alle Wörter, die die Binärdarstellung einer durch 3 teilbaren Zahl darstellen, also

$$0, 11, 110, 1001, 1100, 1111, 10010, \ldots$$

  Hausaufgabe!

Satz 39

Eine Sprache $L \subseteq \Sigma^*$ ist genau dann durch einen regulären Ausdruck darstellbar, wenn sie regulär ist.

Beweis:

"$\Longrightarrow$":

Sei also $L = L(\gamma)$.

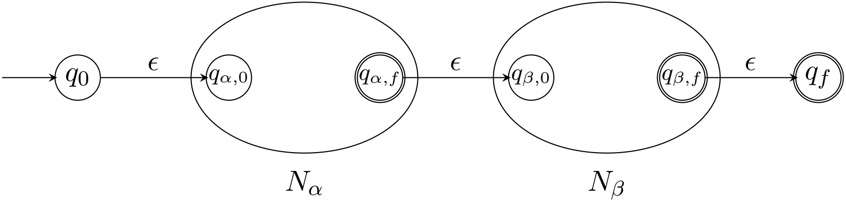
Wir zeigen: $\exists$ NFA N mit $L = L(N)$ mit Hilfe strukureller Induktion.

Induktionsanfang: Falls $\gamma = \emptyset$, $\gamma = \epsilon$, oder $\gamma = a \in \Sigma$, so folgt die Behauptung unmittelbar.

Induktionsschritt:

$\gamma = \alpha\beta$:
nach Induktionsannahme $\exists$ NFA N_α und N_β mit

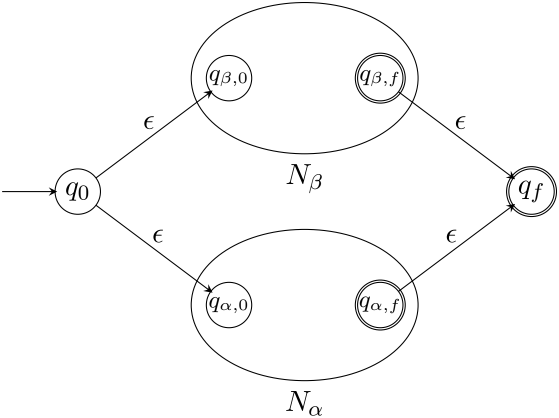
$$L(N_\alpha) = L(\alpha) \text{ und } L(N_\beta) = L(\beta)\ .$$



Induktionsschritt (Forts.):

$\gamma = (\alpha \mid \beta)$:
nach Induktionsannahme $\exists$ NFA N_α und N_β mit

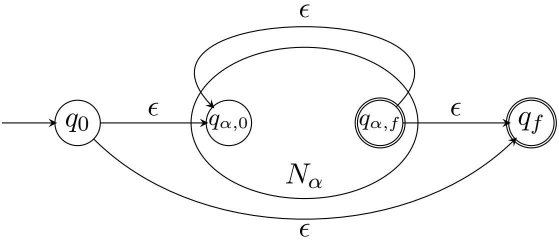
$$L(N_\alpha) = L(\alpha) \text{ und } L(N_\beta) = L(\beta)\ .$$



Induktionsschritt (Forts.):

$\gamma = (\alpha)^*$:
nach Induktionsannahme $\exists$ NFA N_α mit

$$L(N_\alpha) = L(\alpha)\ .$$



"$\Longleftarrow$":

Sei $M = (Q, \Sigma, \delta, q_0, F)$ ein deterministischer endlicher Automat. Wir zeigen: es gibt einen regulären Ausdruck γ mit $L(M) = L(\gamma)$.

Sei $Q = \{q_0, \ldots, q_n\}$. Wir setzen

$$R_{ij}^k := \{w \in \Sigma^*;\ \text{die Eingabe } w \text{ überführt den im Zustand } q_i \text{ gestarteten Automaten in den Zustand } q_j, \text{ wobei alle zwischendurch durchlaufenen Zustände einen Index kleiner gleich } k \text{ haben}\}$$

Behauptung: Für alle $i, j \in \{0, \ldots, n\}$ und alle $k \in \{-1, 0, 1, \ldots, n\}$ gilt:
Es gibt einen regulären Ausdruck α_{ij}^k mit $L(\alpha_{ij}^k) = R_{ij}^k$.

Bew.:

Induktion über k:
$k = -1$: Hier gilt

$$R_{ij}^{-1} := \begin{cases} \{a \in \Sigma;\ \delta(q_i, a) = q_j\}, & \text{falls } i \neq j \\ \{a \in \Sigma;\ \delta(q_i, a) = q_j\} \cup \{\epsilon\}, & \text{falls } i = j \end{cases}$$

R_{ij}^{-1} ist also endlich und lässt sich daher durch einen regulären Ausdruck α_{ij}^{-1} beschreiben.

Bew.:

Induktion über k:
$k \Rightarrow k+1$: Hier gilt

$$R_{ij}^{k+1} = R_{ij}^{k} \cup R_{i\,k+1}^{k}(R_{k+1\,k+1}^{k})^{*}R_{k+1\,j}^{k}$$
$$\alpha_{ij}^{k+1} = (\alpha_{ij}^{k} \mid \alpha_{i\,k+1}^{k}(\alpha_{k+1\,k+1}^{k})^{*}\alpha_{k+1\,j}^{k})$$

Somit gilt: $L(M) = L((\alpha_{0\,f_1}^{n} \mid \alpha_{0\,f_2}^{n} \mid \cdots \mid \alpha_{0\,f_r}^{n}))$, wobei $f_1, \ldots, f_r$ die Indizes der Endzustände seien.

$\square$(Satz 39)

3.8 Abschlusseigenschaften regulärer Sprachen

Satz 40

Seien $R_1, R_2 \subseteq \Sigma^*$ reguläre Sprachen. Dann sind auch

$$R_1 R_2, R_1 \cup R_2, R_1^*, \Sigma^* \setminus R_1 (=: \bar{R}_1), R_1 \cap R_2$$

reguläre Sprachen.

Beweis:

$R_1 R_2, R_1 \cup R_2, R_1^*$ klar.

$\Sigma^* \setminus R_1$	:	Sei $R_1 = L(A), A$ DFA, $A = (Q, \Sigma, \delta, q_0, F)$, δ vollständig. Betrachte $A' = (Q, \Sigma, \delta, q_0, Q \setminus F)$. Dann ist $L(A') = \Sigma^* \setminus L(A)$
$R_1 \cap R_2$	:	De Morgan

$\square$

Definition 41

Substitution (mit regulären Mengen) ist eine Abbildung, die jedem $a \in \Sigma$ eine reguläre Menge $h(a)$ zuordnet. Diese Abbildung wird kanonisch auf Σ^* erweitert.

Ein Homomorphismus ist eine Substitution mit $\forall a \in \Sigma : |h(a)| = 1$

Satz 42

Reguläre Sprachen sind unter (regulärer) Substitution, Homomorphismus und inversem Homomorphismus abgeschlossen.

Beweis:

Wir zeigen (nur) die Behauptung für den inversen Homomorphismus.

Sei $h : \Delta \to \Sigma^*$ ein Homomorphismus, und sei $R \subseteq \Sigma^*$ regulär.

Zu zeigen: $h^{-1}(R) \subseteq \Delta^*$ ist regulär.

Sei $A = (Q, \Sigma, \delta, q_0, F), L(A) = R$.

Betrachte $A' = (Q, \Delta, \delta', q_0, F)$, mit

$$\delta'(q, a) = \delta(q, h(a)) \quad \forall q \in Q, a \in \Delta\ ,$$

wobei wir nunmehr der Einfachheit halber statt $\hat{\delta}$ nur δ schreiben.

Also gilt

$$\delta'(q_0, w) = \delta(q_0, h(w)) \in F \Leftrightarrow h(w) \in R \Leftrightarrow w \in h^{-1}(R)$$

$\square$