Ausarbeitung der Vorlesung Theoretische Informatik I vom 26. Mai 2000

7. Juni 2000

Zusammenfassung

Diese Vorlesungsveranstaltung beschäftigt sich, anschließend an die vorhergehende, zunächst mit der Sternbildung über Sprachen und anschließend mit der Elimination von ε-Übergängen in nichtdeterministischen Automaten.

Inhaltsverzeichnis

Abbildungsverzeichnis

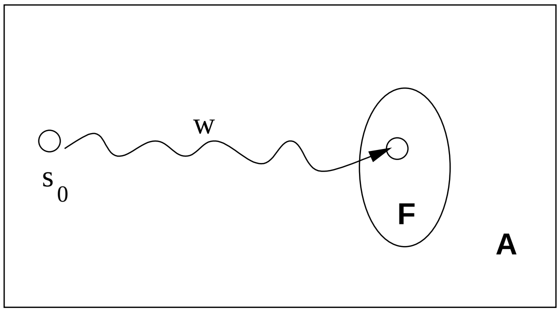


Abbildung 1: Akzeptor für $\mathrm{L(A)}$, $w \in \mathrm{L(A)}$

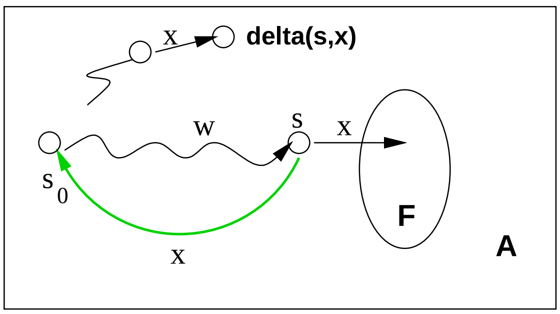


Abbildung 2: Akzeptor für $\mathrm{L(A)}^* \setminus \{\varepsilon\}$, wobei die Teilwörter aus $\mathrm{L(A)}$ und das Gesamtwort aus $\mathrm{L(A)}^*$ stammen

1 Retrospektive

Die vorausgegangene Vorlesung vom 19. Mai 2000 befasste sich mit nichtdeterministischen vollständigen Akzeptoren und deren Äquivalenz zu deterministischen Akzeptoren sowie deren Anwendung zur Erkennung konkatenierter Wörter mittels der sog. *Sternbildung*. Diese Veranstaltung setzt beim entsprechenden Beweis ein.

2 „Sternbildung"

$[\mathrm{L} \in \mathrm{DA} \Rightarrow \mathrm{L}^* \in \mathrm{DA}]$, $R := \mathrm{L(A)}$

2.1 Idee

Die Idee ist die, aus dem Akzeptor aus Abb. 1 den aus Abb. 2 mit gleicher Funktionalität zu konstruieren. Dieser Akzeptor sei nichtdeterministisch. $L^* \ni u = \underbrace{w_1}_{w_1' x} w_2 w_3 \dots \overbrace{w_n}^{w_n' x}$, $w_i \in L$. Die grün eingezeichnete Kante (s, x, s_0) ist die sogenannte *Neustartkante* zur Erkennung mehrerer konkatenierter Wörter. Der Akzeptor erkennt ein einzelnes Teilwort und kehrt, sofern er das Gesamtwort noch nicht vollständig eingelesen hat, über die Neustartkante zu seinem Anfangszustand zurück. Wurde das Gesamtwort komplett eingelesen, so gelangt der Akzeptor in einen Zustand der Endzustandsmenge.

Dieser oben dargestellte spezielle Akzeptor hat eine Besonderheit: Er erkennt das *leere Wort* ε nicht. Dieses gehört jedoch zu $\mathrm{L(A)}^*$, da vereinbart ist, dass $\mathrm{L(A)}^* = \bigcup_{i \in \mathbb{N} \cup \{0\}} \mathrm{L(A)}^i$ auch die

Menge $\mathrm{L(A)}^0 = \{\varepsilon\}$ enthält.[1]

2.2 Behauptung

$$L(B) \setminus \{\varepsilon\} = R^+$$

2.3 Voraussetzung

Sei $R \in \mathrm{DA}$ und $A = (X, S, \delta, s_0, F)$ ein deterministischer endlicher Akzeptor. Wir zeigen: $R^* \in$ DA.

Bilde dazu den nichtdeterministischen Automaten $B = (X, S, \delta', \{s_0\}, F)$ mit

$$\delta' = \begin{cases} \delta(s,x), \text{ falls } \delta(s,x) \notin F \\ \{\delta(s,x), s_0\}, \text{ falls } \delta(s,x) \in F \end{cases}$$

Die erste Zeile der Definition von δ' ergibt sich aus der Überlegung, den Akzeptor so lange in Richtung auf einen Endzustand laufen zu lassen, wie er diesen nicht erreichen würde, das heißt wie er ein Teilwort erkennt (Akzeptor A). Sobald ein Teilwort vollständig erkannt wurde, tritt die Erweiterung durch den Akzeptor B in Kraft: Es wird entweder ein Endzustand angelaufen oder in den Anfangszustand des Akzeptors zurückgewechselt (zweite Zeile der Definition von δ').

2.4 Beweis

Proof. Sei $w = x_1 ... x_r$, $n \geq 1$, $x_i \in X$, $i = 1; ...; r$.

$\boxed{„\subseteq"}$ Sei $w \in L(B)$. Betrachte eine Folge $z_0, z_1, ..., z_r$ mit $z_r \in F$, $z_0 = s_0$, $z_i \in \delta(z_{i-1}, x_i)$ für $i = 1; ...; r$.

Um die jeweiligen „Neustarts" zu suchen, betrachte nun die maximale Teilfolge $z_{i_1}, z_{i_2}, ..., z_{i_k}$ mit $z_{i_j} = s_0$ und $\delta(z_{i_j - 1}, x_{i_j}) \neq z_{i_j}$ für $j = 1; ...; k$. Bilde hierzu das mit jedem Neustart erkannte Teilwort *(Teilwörterbildung)* $w_{i_j} = x_{i_j - 1} \cdots x_{i_j}$ für $j = 1; ...; k$, wobei k die Anzahl der Teilwörter bezeichnet und jedes j einem Neustart des Akzeptors auf dem Weg zur Akzeptanz des Gesamtwortes entspricht.

Anschaulich gesprochen wird folgendes betrachtet: $\underbrace{z_{i_1}}_{=s_0} \overset{w_{i_1}}{\rightsquigarrow} \underbrace{z_{i_2}}_{=s_0} \overset{w_{i_2}}{\rightsquigarrow} \underbrace{z_{i_3}}_{=s_0} \overset{w_{i_3}}{\rightsquigarrow} \cdots \overset{w_{i_{k-1}}}{\rightsquigarrow} \underbrace{z_{i_k}}_{=s_0},$
d.h. auf dem Weg von z_{i_1} nach z_{i_2} wird das Teilwort w_{i_1} vom Akzeptor akzeptiert, es erfolgt ein Neustart und das nächste Teilwort wird akzeptiert etc. Unmittelbar folgt, dass $w = w_{i_1} \cdots w_{i_k}$ und $\delta(s_0, w_{i_j}) \in F$ bezüglich dem Akzeptor A für $j = 1; ...; k$. Dies wiederum bedeutet, dass $w_{i_1} \cdots w_{i_k} \in R$, woraus schlussendlich zu folgern ist, dass gilt: $w \in R^+$.

$\boxed{„\supseteq"}$ Sei $w \in R^+$. Dann gibt es eine Zerlegung $w = w_{i_1} \cdots w_{i_k}$ mit $w_{i_j} \in R$ für $j = 1; ...; k$ (j und k wie oben). Also gilt aufgrund der Randbedingung: $\delta(s_0, w_{i_j}) \in F$ bezüglich dem Akzeptor A. Nach Definition des Akzeptors B folgt: $s_0 \in \delta'(s_0, w_{i_j})$ bezüglich B. Folglich ist $s_0 \in \delta'(s_0, w_{i_1} \cdots w_{i_{k-1}})$ und $\delta'(s_0, w_{i_1} \cdots w_{i_{k-1}} w_{i_k}) \cap F \neq \emptyset$, das bedeutet, der Akzeptor erreicht bei jedem Teilwort, ausgenommen dem letzten, seinen Anfangszustand und nach dem letzten Teilwort einen Endzustand, weshalb der letzte Zustand des Akzeptors mit der Endzustandsmenge F mehr als 0 Elemente (genau eines) gemein hat. Hieraus ist zu folgern, dass $w \in L(B)$, was genau der Behauptung entspricht.

[1]Allgemeine Kleene-Sternbildung: L^*, das heißt der *-Operator auf ein Alphabet L angewandt, bezeichnet die Menge aller Buchstabenketten, die durch Konkatenation von 0 oder mehr Elementen aus L entstehen können. Offensichtlich ergibt die Konkatenation von 0 Elementen das leere Wort ε. (aus: [2])

Da entweder $L(B) = R^*$ oder $L(B) \cup \{\varepsilon\} = R^*$ gelten muss, ist mit $L(B) \in \text{DA}$ auch $R^* \in \text{NVA}$ und somit nach *Satz I*[2] $R \in \text{DA}$ (das ε als Element von R^* wird an dieser Stelle durch das „Hinzuvereinigen" im Falle $L(B) \cup \{\varepsilon\} = R^*$ berücksichtigt).

□

2.5 Fazit

Es gilt also:
deterministische endliche Akzeptoren = nichtdeterministische vollständige Akzeptoren = nichtdeterministische Akzeptoren, oder kurz:

$$\boxed{\text{DNA=NVA=NA}}$$

In einem weiteren Schritt wird den Automaten jetzt die Möglichkeit hinzugefügt, ohne Eingabe eines Zeichens den Zustand zu wechseln. Dies sind die sogenannten

3 Epsilon-Übergänge

3.1 Idee

Lasse spontane Zustandsübergänge ohne Lesen eines Zeichens zu. Dies sind sogenannte *ε-Übergänge*. Diese Übergänge können immer dann stattfinden, wenn das *leere Wort* eingelesen wird. Dieses befindet sich beliebig oft zwischen zwei beliebigen Zeichen eines Wortes: $w = x_1x_2x_3 = x_1\varepsilon ... \varepsilon x_2\varepsilon ... \varepsilon x_3\varepsilon ...$.

3.2 Definition

1. Ein Quintupel $A = (X, S, \delta, s_0, F)$ heißt *nichtdeterministischer endlicher Akzeptor mit ε-Übergängen*, falls gilt:
   - ✘ X, S sind endliche nichtleere Mengen,
   - ✘ $\delta : S \times (X \cup \{\varepsilon\}) \rightarrow 2^S$ ist die Zustandsüberführungsfunktion von A,
   - ✘ $s_0 \subseteq S$ ist die Menge der Anfangszustände und
   - ✘ $F \subseteq S$ ist die Menge der Endzustände.
2. Setze δ wie folgt zu einer Relation $\delta^* : S \times X^* \rightarrow 2^S$ fort: $\forall s \in S, a \in (X \cup \{\varepsilon\}), w \in X^*$ gilt:
   (a) $\delta^*(s, \varepsilon) = \{s\}$
   (b) $\delta^*(s, aw) = \bigcup_{t \in \delta(s,a)} \delta^*(t, w)$.
   δ^* schreiben wir fortan als δ.
3. Die Menge $L(A) = \{w \in X^* | \delta(s_0, w) \cap F \neq \emptyset$ ist die *von A akzeptierte Sprache*.

4 Satz M

Für jeden nichtdeterministischen endlichen Akzeptor A mit ε-Übergängen gilt: $\boxed{L(A) \in \text{DA}}$

[2] Satz I besagt, dass deterministische Akzeptoren äquivalent zu nichtdeterministischen vollständigen Akzeptoren sind, die ihrerseits äquivalent zu nichtdeterministischen Akzeptoren sind. Kurz: DA=NVA=NA. Aufgestellt und bewiesen wurde er in der Vorlesung vom 12. Mai 2000.

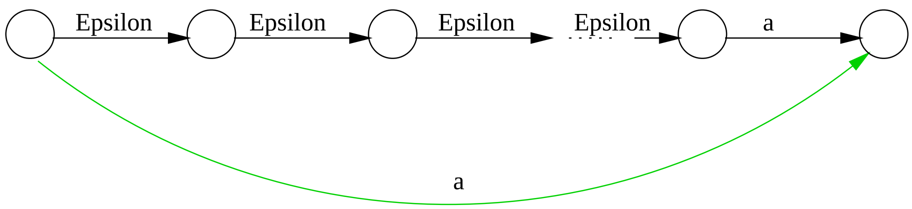


Abbildung 3: Die Elimination von ε-Übergängen

4.1 Beweisidee

Ziehe diejenigen Kanten, die keine ε-Übergänge darstellen, zu Knoten vor, von denen ε-Übergänge zur nächsten Nicht-ε-Kante bestehen. Graphisch veranschaulicht sieht das ganze so aus, wie in Abb. 3 dargestellt. Aus der erkennbaren Kette aus Zuständen und ε-Übergängen wird insgesamt die grüne Kante.

4.2 Beweis

Sei $A = (X, S, \delta, s_0, F)$ ein beliebiger nichtdeterministischer endlicher Akzeptor mit ε-Übergängen. Wir zeigen, dass es einen endlichen nichtdeterministischen Akzeptor A' ohne ε-Übergänge gibt mit $L(A) = L(A')$. Dazu eliminiere die ε-Übergänge in zwei Schritten.

1. Schritt: Elimination von Zyklen Konstruiere für ein geeignetes $n \in \mathbb{N}$ Mengen $S_1, ..., S_n \subseteq S$ mit

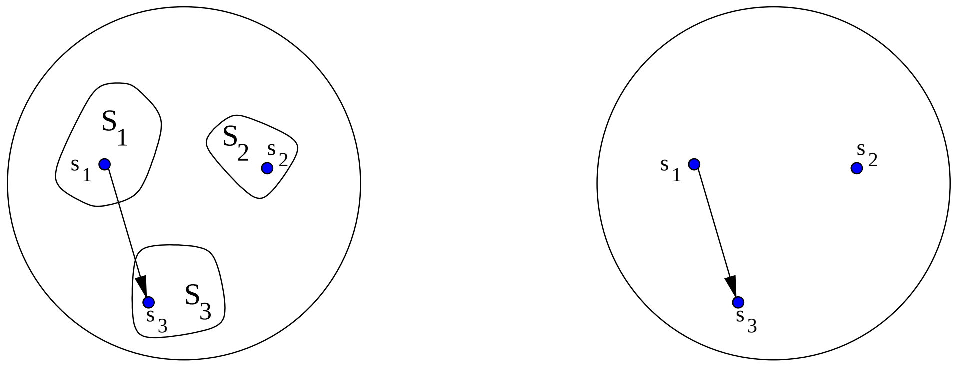
1. $\bigcup\limits_{i=1}^{n} S_i = S$,
2. $\forall i \in \{1; ...; n\}, \forall s, s' \in S_i : [\delta^*(s, \varepsilon) = s']$,
3. $\forall i, j \in \{1; ...; n\}, \forall s \in S_i, \forall s' \in S_j : [i \neq j \Rightarrow s \notin \delta^*(s', \varepsilon) \vee s' \in \delta^*(s, \varepsilon)]$.

Konstruiere dann $\overline{A} = (X, \overline{S}, \overline{\delta}, \overline{s_0}, \overline{F})$ wie folgt:

1. $\overline{S} = \{s_1, ..., s_n\}$; $\overline{s_0} = s_i$, wobei gilt: $s_0 \in S_i$, als Menge der Zustände, wobei Element der Menge $\overline{S}$ die aus den Zustandsmengen S_i des ursprünglichen Automaten extrahierten neuen Einzelzustände sind (zur Verdeutlichung siehe Abb. 4);
2. $\overline{F} = \{s_i | S_i \cap F \neq \emptyset\}$ als Menge der terminalen Zustände,
3. $\forall i, j \in \{1; ...; n\}, \forall a \in X \cup \{\varepsilon\} : [s_i \in \delta(s_j, a) \Leftrightarrow \exists s \in S_i, s' \in S_j : [s \in \delta(s', a) \wedge (i \neq j \vee a = \varepsilon)]]$.

2. Schritt: Elimination verbleibender ε-Übergänge Aus $\overline{A}$ erhält man nun $A' = (X, \overline{S}, \delta', \overline{s_0}, F')$ als neuen Akzeptor mit

1. $F' = \bigcup\limits_{\overline{s} \in \overline{F}} \{\overline{s}' \in \overline{S} | \overline{s} \in \overline{\delta}(\overline{s}', \varepsilon)\}$ und
2. $\forall \overline{s}, \overline{s}' \in \overline{S}, \forall a \in X : [\overline{s}' \in \overline{\delta}(\overline{s}, a) \Leftrightarrow$
   (a) $\overline{s}' \in \overline{\delta}(\overline{s}, a)]$
   (b) $\exists \overline{t} \in \overline{S} : [\overline{s}' \in \overline{\delta}(\overline{t}, a) \wedge \overline{t} \in \overline{\delta}(\overline{s}, \varepsilon)]]$.



Der urspüngliche Automat

Der bearbeitete Automat

Abbildung 4: Elimination von ε-Übergängen aus einem nichtdeterministischen Automaten

Das ist die gesamte Konstruktion des Automaten. Hiermit wurde erreicht, dass Mengen von Zuständen auf einzelne Zustände

zusammengeführt wurden. Graphisch dargestellt wurde aus dem linken Automaten in Abb. 4 der rechte konstruiert.

4.3 Gleichheit der beiden Sprachen

Es gilt: $L(A) = L(A')$.

4.3.1 Beweis

Der Beweis entfällt hier. Es wird auf [1] verwiesen.[3]

5 Vorschau

Die nächste Vorlesung (am 02. Juni 2000) wird sich unter anderem mit Beispielen für die Definition von ε-Übergängen befassen.

Literatur

[1] J. Hopcroft u.a. *Introduction to Automata Theory, Languages and Programming*. Addison-Wesley, 1979.

[2] Katrin Erk u.a. *Theoretische Informatik - Eine umfassende Einführung*. Springer-Verlag, 2000.

[3] Ingo Wegener. *Theoretische Informatik - eine algorithmische Einführung*. B. G. Teubner, Stuttgart, 2. Auflage, 1999.

[3]Die entsprechende Literatur sollte sich in der Universitätsbibliothek unter den Signaturen [ST 130 HOP] (Informatik) bzw. [ES 900 HOP] (Computerlinguistik) befinden, konnte jedoch zum Zeitpunkt der Ausarbeitung dort nicht aufgefunden werden, weshalb an dieser Stelle auf den Beweis nur verwiesen werden kann. Sorry!

Einen Beweis, in welcher Zeit ein A' entsprechender Automat konstruiert werden kann, findet sich übrigens in [3] auf Seite 108f. Ich habe ihn hier nicht aufgeführt, da er nicht unmittelbar mit der oben angesprochenen Situation zusammenhängt.