Einführung in die Computerlinguistik – Satz von Kleene

Dozentin: Wiebke Petersen

5. Foliensatz

Satz von Kleene

(Stephen C. Kleene, 1909 - 1994)

Jede Sprache, die von einem deterministischen endlichen Automaten akzeptiert wird ist regulär und jede reguläre Sprache wird von einem deterministischen endlichen Automaten akzeptiert.

Wiederholung: reguläre Sprachen

reguläre Ausdrücke: Syntax

Die Menge der regulären Ausdrücke RE_Σ über einem Alphabet $\Sigma = \{a_1, \dots, a_n\}$ ist wie folgt definiert:

- $\underline{\emptyset}$ ist ein regulärer Ausdruck.
- ε ist ein regulärer Ausdruck.
- $a_1, \dots, a_n$ sind reguläre Ausdrücke.
- Wenn a und b reguläre Ausdrücke über Σ sind, dann sind
  - $(a + b)$
  - $(a \circ b)$
  - (a^*)

  ebenfalls reguläre Ausdrücke.

Wiederholung: reguläre Sprachen

reguläre Ausdrücke: Semantik

Jeder reguläre Ausdruck r über einem Alphabet Σ beschreibt eine formale Sprache $L(r) \subseteq \Sigma^*$.
Eine formale Sprache ist eine reguläre Sprache, wenn sie durch einen regulären Ausdruck beschrieben werden kann.
Die Funktion L wird induktiv definiert:

- $L(\underline{\emptyset}) = \emptyset$, $L(\varepsilon) = \{\varepsilon\}$, $L(a_i) = \{a_i\}$
- $L(a + b) = L(a) \cup L(b)$
- $L(a \circ b) = L(a) \circ L(b)$
- $L(a^*) = L(a)^*$

Endliche Automaten akzeptieren reguläre Sprachen

Theorem (Kleene)

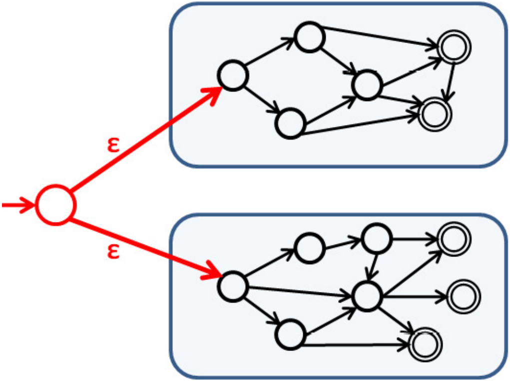
Jede Sprache, die von einem deterministischen endlichen Automaten akzeptiert wird ist regulär und jede reguläre Sprache wird von einem deterministischen endlichen Automaten akzeptiert.

Beweisidee (eine Richtung): Zu jeder regulären Sprache gibt es einen endlichen Automaten, der diese akzeptiert:

$L(\underline{\emptyset})$ $\quad$ $L(\varepsilon)$ $\quad$ $L(a_i)$

Beweis des Satzes von Kleene (Fortsetzung)

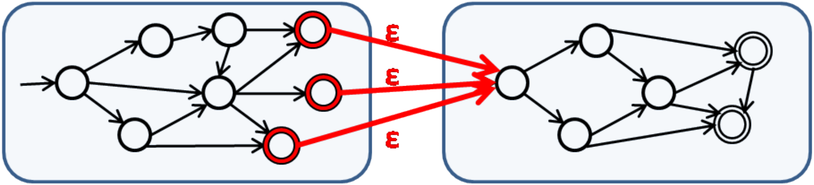
Wenn R_1 und R_2 zwei reguläre Ausdrücke sind und wenn die regulären Sprachen $L(R_1)$ und $L(R_2)$ von den endlichen Automaten $\mathcal{A}_1$ und $\mathcal{A}_2$ akzeptiert werden, dann wird die reguläre Sprache $L(R_1 + R_2)$ von dem folgenden Automaten akzeptiert:



$L(R_1+R_2)$

Beweis des Satzes von Kleene (Fortsetzung)

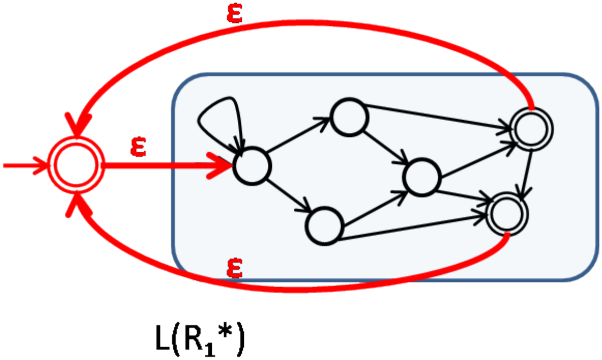
$L(R_1 \circ R_2)$ wird von dem folgenden Automaten akzeptiert:



$L(R_1 \bullet R_2)$

Beweis des Satzes von Kleene (Fortsetzung)

$L(R_1^*)$ wird von dem folgenden Automaten akzeptiert:



Abschlußeigenschaften regulärer Sprachen

Theorem

1. *Wenn L_1 und L_2 reguläre Sprachen sind, dann*
   - *ist die Vereinigung von L_1 und L_2 ($L_1 \cup L_2$) ebenfalls eine reguläre Sprache.*
   - *ist die Schnittmenge von L_1 und L_2 ($L_1 \cap L_2$) ebenfalls eine reguläre Sprache.*
   - *ist die Konkatenation von L_1 und L_2 ($L_1 \circ L_2$) ebenfalls eine reguläre Sprache.*
2. *Das Komplement einer regulären Sprache ist eine reguläre Sprache.*
3. *Wenn L eine reguläre Sprache, dann ist L^* eine reguläre Sprache.*

Beweisidee

a Die Aussage für die Vereinigung, die Konkatenation und den Kleeneschen Stern folgt unmittelbar aus dem Satz von Kleene.

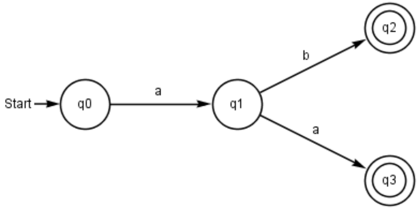
b Die Aussage über den Schnitt zweier regulärer Sprachen folgt aus der Aussage über die Vereinigung, das Komplement und das Gesetz von De Morgan.

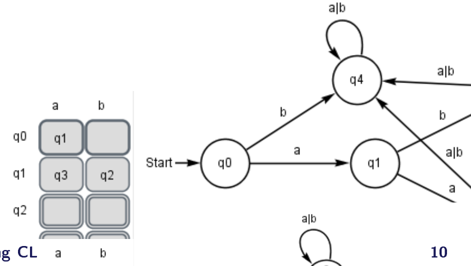
Endlicher Automat zum Komplement einer Sprache

Ein endlicher Automat zum Komplement einer regulären Sprache L wird wie folgt konstruiert:

1. Man starte mit einen deterministischen endlichen Automaten, der L akzeptiert.
2. Man erweitere den Autmaten zu einem deterministischen endlichen Automaten mit vollständiger Übergangsfunktion (Hinzunahme eines "Falle"-Zustands)
3. Man wechsle alle Nichtend- zu

Beispiel $L = \{ab, aa\}$, Alphabet $\Sigma = \{a, b\}$:





Hausaufgabe (BN: entweder Aufgabe 1 oder Aufgabe 2 und 3)

1. Gegeben Seien die beiden regulären Ausdrücke $R_1 = ab^*ab^*$ und $R_2 = a^*b$. Konstruieren Sie für jede der folgenden Sprachen einen endlichen Automaten:
   1. $L(R_1)$, $L(R_2)$
   2. $L(R_1 + R_2)$
   3. $L(R_1 \circ R_2)$ und
   4. $L(R_1^*)$, $L(R_2^*)$
2. Bilden Sie mithilfe des auf den Folien angedeuteten Verfahrens jeweils einen endlichen Automaten, der die folgenden Sprachen akzeptiert:
   Das Komplement im Universum $L((a+b)^*)$ von
   1. $L(ab)$,
   2. $L(ba^* + bb)$
3. Bilden Sie einen Automaten, der die Schnittmenge der beiden Sprachen $L(b^*a^*)$ und $L(ba^+b^*)$ akzeptiert. (Hinweis: Sie müssen nicht stur dem systematischen Verfahren folgen).