Dr. Sebastian Bab WiSe 12/13

Theoretische Grundlagen der Informatik für TI

Termin: VL 11 vom 22.11.2012

Reguläre Ausdrücke

- Reguläre Ausdrücke sind eine lesbarere Notation für Sprachen

  Definition 1 (Regulärer Ausdruck)
  Sei Σ ein Alphabet. Reguläre Ausdrücke sind dann induktiv wie folgt definiert:

  - *$\emptyset$ ist ein regulärer Ausdruck.*
  - *ϵ ist ein regulärer Ausdruck.*
  - *Für alle $a \in \Sigma$ ist a ein regulärer Ausdruck.*
  - *Für zwei reguläre Ausdrücke α, β sind auch $\alpha\beta$, $(\alpha|\beta)$ und $(\alpha)^*$ reguläre Ausdrücke*

- Beispiel mit Alphabet $\Sigma = \{a, b\}$
  - b
  - $(a|abaab)$
  - $(bbb)^*$
  - $(((ab)^*|bba))^*aa$

  Definition 2 (Erzeugte Sprache eines regulären Ausdruckes)
  Sei α ein regulärer Ausdruck. Dann ist $L(\alpha)$, die durch α erzeugte Sprache, wie folgt definiert:

  - *Wenn $\alpha = \emptyset$, dann $L(\alpha) = \emptyset$.*
  - *Wenn $\alpha = \epsilon$, dann $L(\alpha) = \{\lambda\}$.*
  - *Wenn $\alpha = a \in \Sigma$, dann $L(\alpha) = \{a\}$.*
  - *Wenn $\alpha = \beta\gamma$, dann $L(\alpha) = L(\beta)L(\gamma)$.*
  - *Wenn $\alpha = (\beta|\gamma)$, dann $L(\alpha) = L(\beta) \cup L(\gamma)$.*
  - *Wenn $\alpha = (\beta)^*$, dann $L(\alpha) = L(\beta)^*$.*

- fortgeführtes Beispiel

- $L(b) = \{b\}$
- $L((a|abaab)) = \{a, abaab\}$
- $L((bbb)^*) = \{\lambda, bbb, bbbbbb, \dots\}$
- $L((((ab)^*|bba))^*aa) = \{aa, abaa, bbaaa, abbbaabbbaaa, \dots\}$

Satz 3
Jede endliche Sprache ist durch einen regulären Ausdruck beschreibbar.

Beweis:
Sei $L = \{a_1, \dots, a_n\}$ eine beliebige endliche Sprache. Sei

$$\alpha = ((\dots((a_1|a_2)|a_3)\dots|a_{n-1})|a_n)$$

ein regulärer Ausdruck. Dann ist $L(\alpha) = L$. $\square$

Satz 4
Jede durch einen regulären Ausdruck beschriebene Sprache ist regulär.

Beweis:
Wir beweisen den Satz durch strukturelle Induktion über den Aufbau regulärer Ausdrücke.

Induktionsanfang: Sei α ein regulärer Ausdruck der Form

1. $\alpha = \emptyset$. Dann ist $L(\alpha) = L(\emptyset) = \emptyset$. Die leere Sprache ist durch eine reguläre Grammatik ohne Produktionen beschreibbar.
2. $\alpha = \epsilon$. Dann ist $L(\alpha) = L(\epsilon) = \{\lambda\}$. Eine reguläre Grammatik G mit genau einer Produktion
$$S \to \lambda$$
erzeugt genau die gesuchte Sprache.
3. $\alpha = a \in \Sigma$. Dann ist $L(\alpha) = L(a) = \{a\}$. Eine reguläre Grammatik G mit genau einer Produktion
$$S \to a$$
erzeugt genau die gesuchte Sprache.

Induktionsvoraussetzung: Seien α und β zwei reguläre Ausdrücke wobei $L(\alpha)$ und $L(\beta)$ regulär Sprachen sind.

Induktionsschritt: Wir zeigen nun, dass auch $L(\alpha\beta), L((\alpha|\beta))$ und $L((\alpha)^*)$ reguläre Sprachen sind. Wegen der Induktionsvoraussetzung müssen zwei Grammatiken $G_1 = (V_1, \sigma, S_1, P_1)$ und $G_2 = (V_2, \sigma, S_2, P_2)$ existieren mit $L(G_1) = L(\alpha)$ und $L(G_2) = L(\beta)$. Dabei seien V_1 und V_2 disjunkt.

1. zu zeigen: $L(\alpha\beta)$ ist regulär:
   Wir konstruieren eine reguläre Grammatik $G = (V, \Sigma, S, P)$ mit $V = V_1 \cup V_2$ und $S = S_1$ so, dass $L(G) = L(\alpha\beta)$. Die Produktionen setzen sich folgendermaßen zusammen:
   - Falls $\lambda \in L(\alpha)$ und $\lambda \in L(\beta)$, dann gilt $S \to \lambda \in P$.
   - Falls $\lambda \in L(\alpha)$ und $\lambda \notin L(\beta)$, dann gilt $S \to aA_2 \in P$ für alle $S_2 \to aA_2 \in P_2$ und $S \to a \in P$ für alle $S_2 \to a \in P_2$.
   - Für alle $A_{1/2} \to aB_{1/2} \in P_{1/2}$ gilt auch $A_{1/2} \to aB_{1/2} \in P$
   - Für alle $A_1 \to a \in P_1$ gilt auch $A_1 \to aS_2 \in P$
   - Falls $\lambda \in L(\beta)$, dann gilt $A_1 \to a \in P$ für alle $A_1 \to a \in P_1$
   - Für alle $A_2 \to a \in P_2$ gilt auch $A_2 \to a \in P$

   Die angegebene Grammatik arbeitet wie G_1. Wenn ein Wort durch G_1 erzeugt wurde, wechselt sie in den Startzustand von G_2 statt aufzuhören und hängt ein Wort an, welches durch G_2 erzeugt wird.
2. zu zeigen: $L((\alpha|\beta))$ ist regulär:
   Wir konstruieren eine reguläre Grammatik $G = (V, \Sigma, S, P)$ mit $V = V_1 \cup V_2 \cup \{S\}$ und $S \notin V_1$ und $S \notin V_2$ so, dass $L(G) = L((\alpha|\beta))$. Die Produktionen setzen sich folgendermaßen zusammen:
   - Falls $\lambda \in L(\alpha)$ oder $\lambda \in L(\beta)$, dann gilt $S \to \lambda \in P$.
   - Für alle $S_{1/2} \to aA_{1/2} \in P_{1/2}$ ist auch $S \to aA_{1/2} \in P$
   - Für alle $S_{1/2} \to a \in P_{1/2}$ ist auch $S \to a \in P$
   - Für alle $A_{1/2} \to aB_{1/2} \in P_{1/2}$ ist auch $A_{1/2} \to aB_{1/2} \in P$
   - Für alle $A_{1/2} \to a \in P_{1/2}$ ist auch $A_{1/2} \to a \in P$

   Die angegebene Grammatik kann das leere Wort direkt vom Startsymbol her erzeugen, falls dies notwendig ist. Sie entscheidet sich ansonsten direkt am Startsymbol für eine Untergrammatik und arbeitet dann in dieser Grammatik weiter.
3. zu zeigen: $L((\alpha)^*)$ ist regulär:
   Wir konstruieren eine reguläre Grammatik $G = (V, \Sigma, S, P)$ mit $V = V_1 \cup \{S\}$ und $S \notin V_1$ so, dass $L(G) = L((\alpha)^*)$. Die Produktionen setzen sich folgendermaßen zusammen:
   - $S \to \lambda \in P$.
   - Für alle $S_1 \to aA_1 \in P_1$ ist auch $S \to aA_1 \in P$
   - Für alle $S_1 \to a \in P_1$ ist auch $S \to a \in P$
   - Für alle $A_1 \to aB_1 \in P_1$ ist auch $A_1 \to aB_1 \in P$
   - Für alle $A_1 \to a \in P_1$ sind auch $A_1 \to a \in P$ und $A_1 \to aS_1 \in P$

Die angegebene Grammatik kann das leere Wort direkt vom Startsymbol her erzeugen. Sie erzeugt ein Wort genau wie dies in Grammatik G_1 geschieht, mit dem Unterschied, dass an den Stellen an denen G_1 terminiert, die Grammatik G die Wahl hat, ein weiteres Wort anzuhängen, oder ebenfalls zu terminieren.

□

Satz 5
Reguläre Sprachen sind abgeschlossen unter den Operationen $^*, \cup, \circ$. *Also sind für zwei reguläre Sprachen* L_1 *und* L_2 *auch* $L_1^*, L_1 \cup L_2, L_1 \circ L_2$ *reguläre Sprachen.*

Die Beweise verlaufen analog zum Beweis, dass reguläre Ausdrücke reguläre Sprachen erzeugen.

- Den Umgang mit „Rechenregeln"für reguläre Ausdrücke nennt man Reguläre Algebra.

Satz 6
(Kommutativität) Seien α *und* β *zwei reguläre Ausdrücke. Dann gilt*

$$L((\alpha|\beta)) = L((\beta|\alpha)).$$

- Es gilt allerdings nicht $L(\alpha\beta) = L(\beta\alpha)$, wie man an dem einfachen Beispiel $\alpha = a, \beta = b$ erkennen kann. Es würde sich sonst $\{ab\} = \{ba\}$ ergeben, was offensichtlich nicht der Fall ist.

Satz 7
(Assoziativität) Seien α, β *und* γ *reguläre Ausdrücke. Dann gilt*

$$L(((\alpha|\beta)|\gamma)) = L(\alpha|(\beta|\gamma))$$

und

$$L((\alpha\beta)\gamma) = L(\alpha(\beta\gamma)).$$

- Assoziativitätsgesetze weisen oft auf unnötige Klammerung hin. Daher kann man die Klammerung in diesen Fällen meist einfach weglassen.

Satz 8
(Neutrales Element) Sei α *ein regulärer Ausdruck. Dann gilt*

$$L((\alpha|\emptyset)) = L(\emptyset|\alpha) = L(\alpha)$$

und

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

Satz 9
(Annihilator) Sei α ein regulärer Ausdruck. Dann gilt

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

- Da es in der leeren Sprache keine Worte gibt, findet sich auch kein Wort, welches verknüpft wurde aus einem Wort einer beliebigen Sprache und einem Wort aus der leeren Sprache.

Satz 10
(Distributivität) Seien α, β und γ reguläre Ausdrücke. Dann gilt

$$L(\alpha(\beta|\gamma)) = L((\alpha\beta|\alpha\gamma))$$

und

$$L((\alpha|\beta)\gamma) = L((\alpha\gamma|\beta\gamma)).$$

Satz 11
(Idempotenz) Sei α ein regulärer Ausdruck. Dann gilt

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

Satz 12
(Hüllengesetze) Sei α ein regulärer Ausdruck. Dann gilt

- $L(((\alpha)^*)^*) = L((\alpha)^*)$ *(mehrfache Anwendung des Sternoperators liefert keine weiteren Worte),*
- $L((\emptyset)^*) = L(\epsilon)$ *(das leere Wort ist immer in der Kleeneschen Hülle einer Sprache),*
- $L((\epsilon)^*) = L(\epsilon)$

- Die Ausdrücke $\emptyset$ und ϵ sind die einzigen regulären Ausdrücke, die nach Hüllenbildung immernoch endliche Sprache erzeugen.
- Es lassen sich weitere Gesetze, wie z.B. $L((\alpha|\beta)^*) = L(((\alpha)^*(\beta)^*)^*)$ finden.