Grundzüge der Theoretischen Informatik

29. Oktober 2021

Markus Bläser

Universität des Saarlandes

Kapitel 3: Reguläre Ausdrücke

Formale Definition

$((\emptyset + (01))^*)$

$(((\emptyset 0)0)0) \cdots$

Definition (3.1)

Sei Σ ein Alphabet. Die Zeichen "(", ")", "$\underline{\emptyset}$", "$\underline{\varepsilon}$", "$+$", und "$*$" seien nicht in Σ.

Reguläre Ausdrücke über Σ sind wie folgt definiert.

1. $\underline{\emptyset}$ und $\underline{\varepsilon}$ sind reguläre Ausdrücke.
2. σ is a regulärer Ausdruck, $\sigma \in \Sigma$.
3. Sind E und F reguläre Ausdrücke, so auch $(E + F)$, (EF) und (E^*).

Semantik

$$L(\,(((01)0)+\varepsilon)\,)$$

$$= L(((01)0)) \cup L(\varepsilon)$$

Definition (3.2)

Sei E ein regulärer Ausdruck. L(E) ist definiert wie folgt.

1. Falls $E = \emptyset$, dann ist $L(E) = \emptyset$.
   Falls $E = \varepsilon$, dann ist $L(E) = \{\varepsilon\}$.
2. Falls $E = \sigma$ für $\sigma \in \Sigma$, dann ist $L(E) = \{\sigma\}$.
3. Falls $E = (E_1 + E_2)$, dann ist $L(E) = L(E_1) \cup L(E_2)$.
   Falls $E = (E_1 E_2)$, dann ist $L(E) = L(E_1)L(E_2)$.
   Falls $E = (E_1^*)$, dann ist $L(E) = L(E_1)^*$.

$$= L((01))\{0\} \cup \{\varepsilon\}$$

$$= L(0)L(1)\{0\} \cup \{\varepsilon\}$$

$$= \{0\}\{1\}\{0\} \cup \{\varepsilon\} = \{010\} \cup \{\varepsilon\}$$

Rechengesetze

Wir schreiben $E = F$, falls $L(E) = L(F)$.

Theorem (3.3)

Für alle regulären Ausdrücke E, F *und* G *gilt:*

1. $E + F = F + E$ *(Kommutativität von* $+$*),*
2. $(E + F) + G = E + (F + G)$ *(Assoziativität von* $+$*),*
3. $(EF)G = E(FG)$ *(Assoziativität von Konkatenation),*
4. $\emptyset + E = E + \emptyset = E$ *(*$\emptyset$ *ist neutrales Element bzgl.* $+$*),*
5. $\varepsilon E = E\varepsilon = E$ *(*ε *ist neutrales Element bzgl. Konkatenation),*
6. $\emptyset E = E\emptyset = \emptyset$ *(*$\emptyset$ *Nullelement bzgl. Konkatenation)*
7. $E + E = E$ *(Vereinigung ist idempotent),*
8. $(E + F)G = (EG) + (FG)$ *(Rechts-Distributivgesetz),*
9. $E(F + G) = (EF) + (EG)$ *(Links-Distributivgesetz),*
10. $(E^*)^* = E^*$,
11. $\emptyset^* = \varepsilon$,
12. $\varepsilon^* = \varepsilon$.

$3+4=4+3$

als Strings unterschiedlich

aber beschreiben die gleiche Zahl

Beweis 3.3.

1) $L(E+F) = L(E) \cup L(F)$

$= L(F) \cup L(E)$

$= L(F+E)$

5) $L(\varepsilon E) = \{\varepsilon\} L(E)$

$= \{ ab \mid a \in \{\varepsilon\} \wedge b \in L(E) \}$

$= \{ \underbrace{\varepsilon b}_{b} \mid b \in L(E) \}$

$= L(E)$

$L(E\varepsilon)$ analog

10) $L := L(E)$

$L^* \overset{!}{=} (L^*)^*$

$L^* \subseteq (L^*)^*$ ✓

$(L^*)^* \subseteq L^*$

Seien $x_1, \dots, x_n \in L^*$

Dann gibt es $y_{\nu,1}, \dots, y_{\nu,j_\nu} \in L$

mit $x_\nu = y_{\nu,1} \dots y_{\nu,j_\nu}$, $1 \leq \nu \leq n$

$x_1 x_2 \dots x_n = y_{11} \dots y_{1j_1} y_{21} \dots y_{2j_2}$

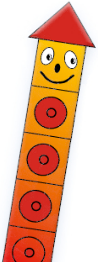
$\dots y_{n,1} \dots y_{nj_n} \in L^*$

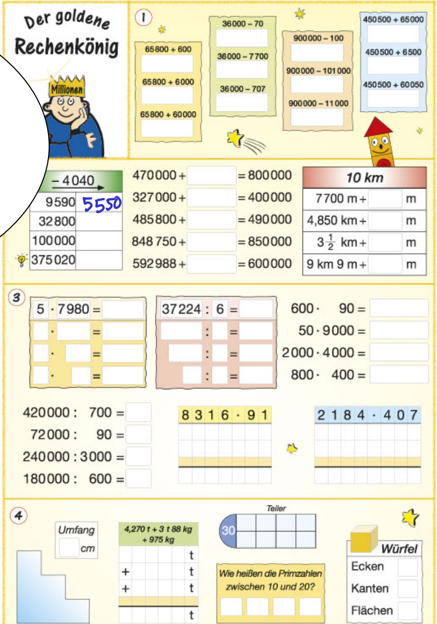
$\Rightarrow (L^*)^* \subseteq L^*$

Vorfahrtsregeln

$((ab)+c) \rightsquigarrow ab+c$

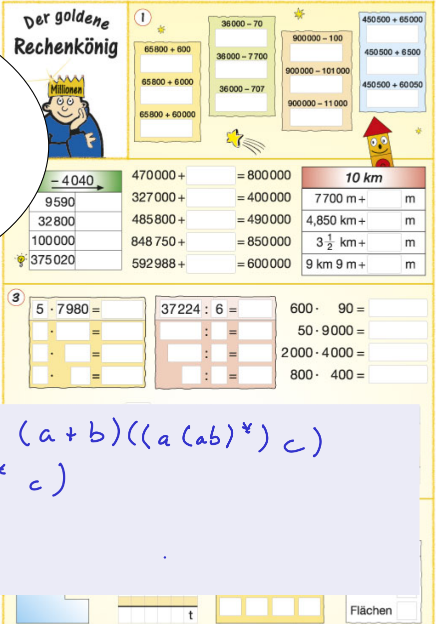
1. Klammern zuerst!
2. Kleensche Hülle $*$ vor Konkatenation.
3. Konkatenation vor Vereinigung $+$.

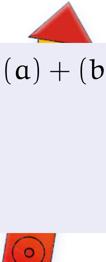




Vorfahrtsregeln

1. Klammern zuerst!
2. Kleensche Hülle * vor Konkatenation.
3. Konkatenation vor Vereinigung $+$.





$$((a)+(b))(((a)((a)(b))^*)(c)) = (a+b)((a(ab)^*)c)$$

$$= (a+b)(a(ab)^*c)$$

```
> more obstsalat.txt
```

```
> more obstsalat.txt
banane
apfel
birne
ananas
kiwi
```

```
> more obstsalat.txt
banane
apfel
birne
ananas
kiwi
> egrep '(an){2}' obstsalat.txt
```

```
> more obstsalat.txt
banane
apfel
birne
ananas
kiwi
> egrep '(an){2}' obstsalat.txt
banane
ananas
```

```
> more obstsalat.txt
banane
apfel
birne
ananas
kiwi
> egrep '(an){2}' obstsalat.txt
banane
ananas
> egrep '(k|b)i' obstsalat.txt
```

anan

(k+b) i

```
> more obstsalat.txt
banane
apfel
birne
ananas
kiwi
> egrep '(an){2}' obstsalat.txt
banane
ananas
> egrep '(k|b)i' obstsalat.txt
birne
kiwi
```

Reguläre Ausdrücke und endliche Automaten

Theorem (3.5)

Falls E *ein regulärer Ausdruck ist, dann ist* $L(E) \in \mathrm{REG}$.

Theorem (3.6)

Für jeden deterministischen endlichen Automaten $M = (Q, \Sigma, \delta, q_0, Q_{acc})$ *gibt es einen regulären Ausdruck* E *mit* $L(M) = L(E)$.

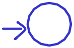
Beweis 3.5

„ E regulärer Ausdruck $\Rightarrow$ L(E) regulär "

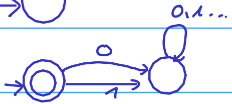
strukturelle Induktion

Induktionsanfang

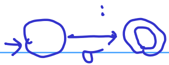
1) $E = \underline{\emptyset}$, $L(\underline{\emptyset}) = \emptyset$



2) $E = \underline{\varepsilon}$, $L(\underline{\varepsilon}) = \{\varepsilon\}$



3) $E = \sigma$, $L(\underline{\sigma}) = \{\sigma\}$



Induktionsschritt

Sei 1) $E = E_1 \cup E_2$ oder 2) $E = E_1 E_2$ oder 3) $E = E_1^*$

Indvor: $L(E_1)$ und $L(E_2)$ regulär

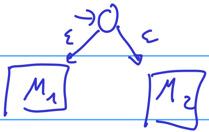
z.z: $L(E)$ ist regulär

1) $L(E) = L(E_1) \cup L(E_2)$

REG ist unter Vereinigung abgeschlossen.

a) Produktautomat

b)



2) $L(E) = L(E_1)L(E_2)$

$L(E)$ ist regulär, da REG unter Konkatenation abgeschlossen ist

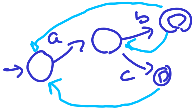
3) $L(E) = L(E_1)^*$

REG ist unter Kleenescher Hülle abgeschlossen. $\square$

$$A^* = \{ a \mid \exists m \in \mathbb{N} \wedge \exists x_1, x_2, \dots, x_m \in A : a = x_1 \dots x_m \}$$

$$\{a\}^* = \{ \varepsilon, a, aa, aaa, \dots \}$$

Beweis von Theorem 3.6



$a(b+c)$

$M = (Q, \Sigma, \delta, q_0, Q_{acc})$

O.E. $Q = \{1, 2, \dots, n\}$, $q_0 = 1$, $Q_{acc} = \{j_1, \dots, j_e\}$

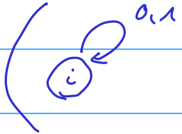
Wir konstruieren Ausdrücke E_{ij}^{k} so dass

$$L(E_{ij}^{k}) = \{ w \mid \delta^*(i, w) = j \text{ und für jeden Präfix } w' \text{ von } w \text{ mit } w' \neq \varepsilon \text{ und } w' \neq w \text{ gilt } \delta^*(i, w') \leq k \}$$

Induktion in k

$k = 0$: M darf keinen anderen Zustand besuchen, wenn er von i nach j geht.

$E_{ij}^0 = \sigma_1 + .. + \sigma_t$, wobei $\sigma_1, .., \sigma_t$ die Kantenbeschriftungen der Kanten von i nach j sind. $(i \neq j)$

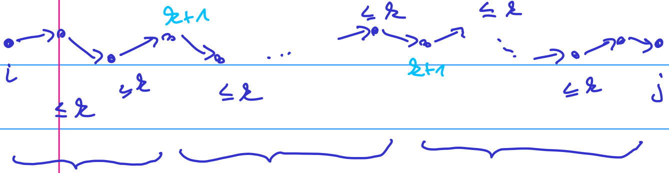


$E_{ii}^0 = \varepsilon + \sigma_1 + .. + \sigma_t$

wobei, ...

Induktionsschritt: $k \to k+1$

Induktionsannahme: E_{ij}^{k} sind konstruiert, $1 \le i,j \le n$



$E^{k}_{i,k+1}$ $\quad$ $E^{k}_{k+1,k+1}$ $\quad$ $E^{k}_{k+1\,j}$

$$E_{ij}^{k+1} := E^{k}_{i,k+1} \left(E^{k}_{k+1,k+1} \right)^* E^{k}_{k+1\,j} + E^{k}_{ij}$$

Formal bleibt zu zeigen: Wenn ein Wort w „die Eigenschaft" hat, dann ist $w \in L(E_{ij}^{k+1})$ □

$$L(M) = L(E^{n}_{1,j_1} + \dots + E^{n}_{1,j_l}) \qquad \square$$