

Grundzüge der Theoretischen Informatik

29. Oktober 2021

Markus Bläser

Universität des Saarlandes

Kapitel 3: Reguläre Ausdrücke

Formale Definition

$$((\sigma + (\sigma))^{*})$$

$$(((\sigma\sigma)\sigma)) \dots$$

Definition (3.1)

Sei Σ ein Alphabet. Die Zeichen “(”, “)”, “ \emptyset ”, “ ε ”, “+”, und “*” seien nicht in Σ .

Reguläre Ausdrücke über Σ sind wie folgt definiert.

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

$$\begin{aligned} L(((01)0) + \varepsilon) \\ = L((01)0) \cup L(\varepsilon) \end{aligned}$$

Definition (3.2)

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

1. Falls $E = \emptyset$, dann ist $L(E) = \emptyset$.
Falls $E = \underline{\varepsilon}$, 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 Konkatination),
4. $\emptyset + E = E + \emptyset = E$ (\emptyset ist neutrales Element bzgl. $+$),
5. $\varepsilon E = E\varepsilon = E$ (ε ist neutrales Element bzgl. Konkatination),
6. $\emptyset E = E\emptyset = \emptyset$ (\emptyset Nullelement bzgl. Konkatination)
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 Summanden unterschiedlich

aber beschreiben die gleiche Zahl

Beweis 3.3.

$$\begin{aligned} 1) \quad L(E + F) &= L(E) \cup L(F) \\ &= L(F) \cup L(E) \\ &= L(F + E) \end{aligned}$$

$$\begin{aligned} 5) \quad L(\varepsilon E) &= \{\varepsilon\} L(E) \\ &= \{ab \mid a \in \{\varepsilon\} \wedge b \in L(E)\} \\ &= \{\underbrace{\varepsilon}_b b \mid b \in L(E)\} \\ &= L(E) \end{aligned}$$

$L(E_c)$ analog

$$10) \quad L := L(E)$$

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

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

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

$$\text{Seien } x_1, \dots, x_n \in L^*$$

$$\text{Dann gibt es } y_{v,1}, \dots, y_{v,j_v} \in L$$

$$\text{mit } x_v = y_{v,1} \dots y_{v,j_v}, \quad 1 \leq v \leq n$$

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

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

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

Vorfahrtsregeln

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

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

Der goldene Rechenkönig

1

65800 + 600	36000 - 70	900000 - 100	450500 + 65000
65800 + 6000	36000 - 7700	900000 - 101000	450500 + 6500
65800 + 60000	36000 - 707	900000 - 11000	450500 + 60050

2

- 4040	470000 +	= 800000	10 km
9590	327000 +	= 400000	7700 m +
32800	485800 +	= 490000	4,850 km +
100000	848750 +	= 850000	3 $\frac{1}{2}$ km +
375020	592988 +	= 600000	9 km 9 m +

3

5 · 7980 =	37224 : 6 =	600 · 90 =
· =	· =	50 · 9000 =
· =	· =	2000 · 4000 =
· =	· =	800 · 400 =

420000 : 700 =

72000 : 90 =

240000 : 3000 =

180000 : 600 =

8 3 1 6 · 9 1

2 1 8 4 · 4 0 7

4

Umfang cm

4,270 t + 3 t 88 kg + 975 kg

Wie heißen die Primzahlen zwischen 10 und 20?

Würfel

Ecken

Kanten

Flächen

Vorfahrtsregeln

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

Der goldene Rechenkönig

1

65800 + 600
65800 + 6000
65800 + 60000

36000 - 70
36000 - 7700
36000 - 707
36000 - 707

900000 - 100
900000 - 101000
900000 - 11000

450500 + 65000
450500 + 6500
450500 + 60050

2

- 4040
9590
32800
100000
375020

470000 +
327000 +
485800 +
848750 +
592988 +

= 800000
= 400000
= 490000
= 850000
= 600000

3

5 · 7980 =
· =
· =
· =

37224 : 6 =
: =
: =
: =

600 · 90 =
50 · 9000 =
2000 · 4000 =
800 · 400 =

10 km

7700 m + m
4,850 km + m
3 $\frac{1}{2}$ km + m
9 km 9 m + m

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

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


```
> 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

anah

```
> egrep '(an){2}' obstsalat.txt
```

banane

ananas

$$(k+b) i$$

```
> egrep '(k|b)i' obstsalat.txt
```



```
> 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 $L(E) \in \text{REG}$.

Theorem (3.6)

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

Beweis 3.5

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

struktureller Induktion

Induktionsanfang

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



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



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



Induktionsschritt

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

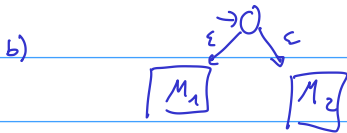
Induktiv: $L(E_1)$ und $L(E_2)$ regulär

z.z.: $L(E)$ ist regulär

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

REG ist unter Vereinigung abgeschlossen.

a) Produktautomat



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

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

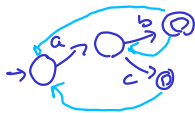
$$3) L(E) = L(E_1)^*$$

REG ist unter Kleinscher 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



$$a(b+c)$$

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

$$\text{O.}\Sigma. \quad Q = \{1, 2, \dots, n\}, \quad q_0 = 1, \quad Q_{acc} = \{j_1, \dots, j_\ell\}$$

Wir konstruieren Ausdrücke E_{ij}^k so dass

$$L(E_{ij}^k) = \{ w \mid \begin{array}{l} \delta^*(i, w) = j \text{ und f\"ur jeden} \\ \text{Pr\"afixe } w' \text{ von } w \text{ mit} \\ w' \neq \varepsilon \text{ und } w' \neq w \text{ gilt} \\ \delta^*(i, w') \leq k \end{array} \}$$

Induktion in k

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

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

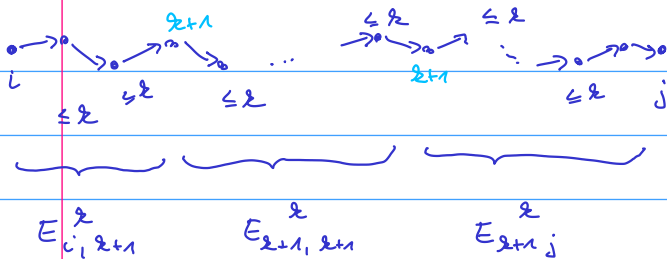
$$E_{ii}^0 = \varepsilon + \sigma_1 + \dots + \sigma_t$$

wobei ε, \dots



Induktionsschritt: $k \rightarrow k+1$

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



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

Formal bleibt zu zeigen: Jeder Weg w

"der Eigenschaft hat, dass mit $w \in \mathcal{L}(E_{ij}^{k+1})$ //

$$L(M) = L(E_{1,j_1}^n + \dots + E_{1,j_\ell}^n)$$

□