Alphabet, Wörter

Endliches Alphabet $\Sigma = \{a, b, c, \dots, z, A, \dots, ␣, \dots\}$

$$\Sigma = \{0,1\}$$

Wort $x = x_1 x_2 \dots x_n$ $x = \text{abrakadabra}$

eine Folge von Buchstaben $x_i \in \Sigma$ $x = 01001$

$n = |x| =$ Länge des Wortes $n=0$: $x = \varepsilon$ (leeres Wort)

$\Sigma^* =$ Menge aller Wörter über dem Alphabet Σ.

Hintereinanderschreiben von Wörter (Verknüpfung, Konkatenation)

$x = abc$, $y = cadb$ $\underset{\text{Wörter}}{xy} = \underset{\text{Buchstaben}}{abccadb} = x \cdot y$

$z = baa$

$xyz = (xy)z = x(yz)$ Verknüpfung ist assoziativ.

nicht kommutativ ($xy \neq yx$ im Allg.)

$(\Sigma^*, \cdot, \varepsilon)$ ist ein Monoid (Halbgruppe mit Einselement).

$$x \cdot \varepsilon = \varepsilon \cdot x = x$$

Potenz $x^i = \underbrace{x \cdot x \cdot x \cdots x}_{i\text{-mal}}$, $x^0 = \varepsilon$, $(01)^{1000}$

Formale Sprachen

<u>Definition</u>: Eine formale Sprache L ist eine Teilmenge von Σ^*.

<u>Bsp.</u> (Syntaktisch) gültige Programme in JAVA.
- (Grammatikalisch korrekte) deutsche Sätze. ??
- $L = \{ x \in \{0,1\}^* \mid x \text{ enthält genauso viele Nullen wie Einsen} \}$
- $L = \emptyset$, $L = \Sigma^*$
- $L = \{2,3,5,7,11,13, \dots\} = \{\text{Dezimaldarstellungen aller Primzahlen}\} \in \{0,1,\dots,9\}^*$
- $L = \{ \text{JAVA-Programme, die keine Eingabe lesen und unendlich lange laufen} \}$

Jedes algorithmische Problem mit einer JA-NEIN-Ausgabe (Entscheidungsproblem) führt zu einer formalen Sprache.
- $L = \{ \text{Menge der Eingaben mit Antwort JA} \}$

Operationen auf Sprachen:
- Mengenoperationen $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 \setminus L_2$
- Hintereinanderschreiben:

$$L_1 L_2 = L_1 \cdot L_2 = \{ xy \mid x \in L_1, y \in L_2 \}$$

$$\{aba, ab\} \cdot \{a, aa\} = \{\underline{ab}aa, abaaa, ab.a, a\underline{b}aa\}$$

$$\{abaa, abaaa, aba\}$$

Hintereinanderschreiben von Sprachen ist ebenfalls assoziativ.

$$(L_1 L_2) L_3 = L_1 (L_2 L_3)$$

$(\mathcal{P}(\Sigma^*), \cdot, ?)$ ist auch ein Monoid.

Potenzmenge

Potenz $\quad L^i = \underbrace{L \cdot L \cdots L}_{i\text{-mal}} \qquad L^0 = \{\varepsilon\}$

$$\boxed{L^* = \bigcup_{i \geq 0} L^i = \underset{=\{\varepsilon\}}{L^0} \cup L \cup L.L \cup L.L.L \cup \cdots}$$

$\varepsilon \in L^*$

$\Sigma^i = \{\text{Wörter der Länge } i\}, \quad \Sigma^* = \Sigma^*$

$\{\varepsilon, -, +\}\{1, 2, \ldots, 9\}\{0, 1, 2, \ldots, 9\}^* \{\boxed{,}\}\{0, 1, \ldots, 9\}\{0, 1, \ldots, 9\}^*$

3,14159 $\quad$ 7,22 $\quad$ 305,6 $\quad$ ~~01,01~~ $\quad$ ~~,7~~ $\quad$ ~~3,~~ $\quad$ ~~0,5~~

$L^+ = \bigcup_{i \geq 1} L^i \quad$ (mindestens eine Wiederholung)

$$\left(\{\varepsilon,-,+\}\{1,2,\dots,9\}\{0,1,2,\dots,9\}^* \cup \{0\}\right)\left(\{\boxed{,}\}\{0,1,\dots,9\}^+ \cup \{\varepsilon\}\right)$$

Ein *regulärer Ausdruck* ist ein Ausdruck, der eine Sprache, ausgehend von endlichen Sprachen, mit Hilfe der Operationen $\cup$, $\cdot$, $*$ darstellt.

Es gibt verschiedene Konventionen, wie man reguläre Ausdrücke kodieren kann.

Problem: Unterscheidung zwischen Σ und Metazeichen

Eine *reguläre Sprache* ist eine Sprache, die durch einen regulären Ausdruck beschrieben werden kann.