Grundbegriffe der Informatik

Einheit 5: formale Sprachen

Prof. Dr. Tanja Schultz

Karlsruher Institut für Technologie, Fakultät für Informatik

Wintersemester 2012/2013

Überblick

Sprachen

- Natürliche Sprachen sind gekennzeichnet durch:
  - Aussprache
  - Stil, z. B. Wortwahl und Satzbau
  - Welche Formulierungen sind syntaktisch korrekt?
  - Ist und syntaktisch welcher welcher Satz korrekt nicht?
- Informatik:
  - Sprachen, die nicht natürlich sind:
    - Programmiersprachen
    - Aufbau von Emails, WWW-Seiten, ...
    - Eingabedateien für ...
  - Syntax
    - Wie spezifiziert man, was korrekt ist?
    - Wie überprüft man, ob etwas korrekt ist?
  - Semantik
    - Wie definiert man, was syntaktisch korrekte Gebilde bedeuten?
    - Darum kümmern wir uns später

Überblick

Formale Sprachen

- Alphabet A gegeben
- Eine formale Sprache (über einem Alphabet A) ist eine Teilmenge $L \subseteq A^*$.
- Im Zusammenhang mit syntaktischer Korrektheit:
  - formale Sprache L der syntaktisch korrekten Gebilde
  - syntaktisch falsche Gebilde gehören *nicht* zu L

Beispiele:

- $A = \{\texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{-}\}$.
  formale Sprache der Dezimaldarstellungen ganzer Zahlen
  - enthält z.B. `1`, `-22` und `192837465`,
  - enthält aber nicht `2-3---41`.
- formale Sprache der syntaktisch korrekten Java-Programme
  - enthält alle Java-Programme
  - enthält aber zum Beispiel *nicht*: `[2] class int)(`

Überblick

Produkt oder Konkatenation formaler Sprachen

- Wir kennen schon die Konkatenation von Wörtern
- Wir definieren das Produkt der Sprachen L_1 und L_2:

$$L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \land w_2 \in L_2\}$$

- Wegen der Assoziativität der Konkatenation von Wörtern $(w_1 \cdot w_2) \cdot w_3 = w_1 \cdot (w_2 \cdot w_3)$ gilt auch:

$$L_1 \cdot L_2 \cdot L_3 = \{w_1 w_2 w_3 \mid w_1 \in L_1 \land w_2 \in L_2 \land w_3 \in L_3\}$$

usw.

- den Punkt „·“ lässt man auch gerne wieder weg

Produkt oder Konkatenation formaler Sprachen

- Wir kennen schon die Konkatenation von Wörtern
- Wir definieren das Produkt der Sprachen L_1 und L_2:

$$L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \land w_2 \in L_2\}$$

- Wegen der Assoziativität der Konkatenation von Wörtern $(w_1 \cdot w_2) \cdot w_3 = w_1 \cdot (w_2 \cdot w_3)$ gilt auch:

$$L_1 \cdot L_2 \cdot L_3 = \{w_1 w_2 w_3 \mid w_1 \in L_1 \land w_2 \in L_2 \land w_3 \in L_3\}$$

usw.

- den Punkt „$\cdot$“ lässt man auch gerne wieder weg

Produkte formaler Sprachen: Beispiele

- wenn $L_1 = \{\mathtt{a}, \mathtt{aa}\}$ und $L_2 = \{\mathtt{b}, \mathtt{ab}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}, \mathtt{aa}\} \cdot \{\mathtt{b}, \mathtt{ab}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}, \mathtt{a} \cdot \mathtt{ab}, \mathtt{aa} \cdot \mathtt{b}, \mathtt{aa} \cdot \mathtt{ab}\} \\ &= \{\mathtt{ab}, \mathtt{aab}, \mathtt{aaab}\} \end{aligned}$$
- wenn $S = \{\mathtt{int}, \mathtt{double}, \mathtt{char}, \dots\}$, $B = \{\mathtt{a}, \dots, \mathtt{z}\}$ und $Z = \{\mathtt{0}, \dots, \mathtt{9}\}$
  dann enthält $S \cdot \{\textvisiblespace\} \cdot B \cdot (B \cup Z)^* \cdot \{\mathtt{;}\}$
  „Deklarationen“ wie z. B.
  - `int␣x42;`
  - `double␣wurzelzwei;`

  aber leider auch
  - `int␣double;`

  andererseits aber nicht
  - `char␣␣hugo␣;`

Neutrales Element des Produkts formaler Sprachen

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \, .$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.
$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Neutrales Element des Produkts formaler Sprachen

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \,.$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.
$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \,.$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.
$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Neutrales Element des Produkts formaler Sprachen

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \,.$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.

$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \,.$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \land w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \land w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.
$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Neutrales Element des Produkts formaler Sprachen

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L \, .$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.

$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Neutrales Element des Produkts formaler Sprachen

Lemma. Für jede formale Sprache L ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L\ .$$

Beweis durch einfaches Nachrechnen:

$$\begin{aligned}
L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \wedge w_2 \in \{\varepsilon\}\} \\
&= \{w_1 w_2 \mid w_1 \in L \wedge w_2 = \varepsilon\} \\
&= \{w_1 \varepsilon \mid w_1 \in L\} \\
&= \{w_1 \mid w_1 \in L\} \\
&= L
\end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$.
$\{\varepsilon\}$ ist das neutrale Element bzgl. des Produkts formaler Sprachen.

Potenzen von Sprachen

- wir wollen Potenzen L^k formaler Sprachen definieren
- „Problem“: Was soll L^0 sein?
- Definiere:

$$L^0 = \{\varepsilon\}$$
$$\forall k \in \mathbb{N}_0 : \; L^{k+1} = L \cdot L^k$$

- Einfaches Nachrechnen ergibt z. B.:

$$L^1 = L$$
$$L^2 = L \cdot L$$
$$L^3 = L \cdot L \cdot L$$

- Genau genommen: $L^3 = L \cdot (L \cdot L)$, aber: Konkatenation von Sprachen ist eine assoziative Operation.

Potenzen von Sprachen

- wir wollen Potenzen L^k formaler Sprachen definieren
- „Problem": Was soll L^0 sein?
- Definiere:

$$\begin{aligned} L^0 &= \{\varepsilon\} \\ \forall k \in \mathbb{N}_0 : \; L^{k+1} &= L \cdot L^k \end{aligned}$$

- Einfaches Nachrechnen ergibt z. B.:

$$\begin{aligned} L^1 &= L \\ L^2 &= L \cdot L \\ L^3 &= L \cdot L \cdot L \end{aligned}$$

- Genau genommen: $L^3 = L \cdot (L \cdot L)$, aber:
  Konkatenation von Sprachen ist eine assoziative Operation.

Potenzen von Sprachen

- wir wollen Potenzen L^k formaler Sprachen definieren
- „Problem“: Was soll L^0 sein?
- Definiere:

$$L^0 = \{\varepsilon\}$$
$$\forall k \in \mathbb{N}_0 : \; L^{k+1} = L \cdot L^k$$

- Einfaches Nachrechnen ergibt z. B.:

$$L^1 = L$$
$$L^2 = L \cdot L$$
$$L^3 = L \cdot L \cdot L$$

- Genau genommen: $L^3 = L \cdot (L \cdot L)$, aber:
  Konkatenation von Sprachen ist eine assoziative Operation.

Potenzen von Sprachen

- wir wollen Potenzen L^k formaler Sprachen definieren
- „Problem“: Was soll L^0 sein?
- Definiere:

$$\begin{aligned} L^0 &= \{\varepsilon\} \\ \forall k \in \mathbb{N}_0 : \; L^{k+1} &= L \cdot L^k \end{aligned}$$

- Einfaches Nachrechnen ergibt z. B.:

$$\begin{aligned} L^1 &= L \\ L^2 &= L \cdot L \\ L^3 &= L \cdot L \cdot L \end{aligned}$$

- Genau genommen: $L^3 = L \cdot (L \cdot L)$, aber:
  Konkatenation von Sprachen ist eine assoziative Operation.

Beispiele für Potenzen von Sprachen (1)

- $L = \{\mathtt{aa}, \mathtt{b}\}$
- Dann ist

$$
\begin{aligned}
L^0 &= \{\varepsilon\} \\
L^1 &= \{\mathtt{aa}, \mathtt{b}\} \\
L^2 &= \{\mathtt{aa}, \mathtt{b}\} \cdot \{\mathtt{aa}, \mathtt{b}\} = \{\mathtt{aa} \cdot \mathtt{aa}, \mathtt{aa} \cdot \mathtt{b}, \mathtt{b} \cdot \mathtt{aa}, \mathtt{b} \cdot \mathtt{b}\} \\
&= \{\mathtt{aaaa}, \mathtt{aab}, \mathtt{baa}, \mathtt{bb}\} \\
L^3 &= \{\mathtt{aa} \cdot \mathtt{aa} \cdot \mathtt{aa},\ \mathtt{aa} \cdot \mathtt{aa} \cdot \mathtt{b},\ \mathtt{aa} \cdot \mathtt{b} \cdot \mathtt{aa},\ \mathtt{aa} \cdot \mathtt{b} \cdot \mathtt{b}, \\
&\qquad \mathtt{b} \cdot \mathtt{aa} \cdot \mathtt{aa},\ \mathtt{b} \cdot \mathtt{aa} \cdot \mathtt{b},\ \mathtt{b} \cdot \mathtt{b} \cdot \mathtt{aa},\ \mathtt{b} \cdot \mathtt{b} \cdot \mathtt{b}\} \\
&= \{\mathtt{aaaaaa}, \mathtt{aaaab}, \mathtt{aabaa}, \mathtt{aabb}, \mathtt{baaaa}, \mathtt{baab}, \mathtt{bbaa}, \mathtt{bbb}\}
\end{aligned}
$$

- Sei
$$L = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}_+\} ,$$
also sozusagen (immer diese Pünktchen ...)
$$L = \{\mathtt{ab}, \mathtt{aabb}, \mathtt{aaabbb}, \mathtt{aaaabbbb}, \dots\} .$$
- Was ist $L^2 = L \cdot L$?
  -
$$\begin{aligned} L^2 = \quad & \{\mathtt{ab} \cdot \mathtt{ab}, \mathtt{ab} \cdot \mathtt{aabb}, \mathtt{ab} \cdot \mathtt{aaabbb}, \dots\} \\ \cup\ & \{\mathtt{aabb} \cdot \mathtt{ab}, \mathtt{aabb} \cdot \mathtt{aabb}, \mathtt{aabb} \cdot \mathtt{aaabbb}, \dots\} \\ \cup\ & \{\mathtt{aaabbb} \cdot \mathtt{ab}, \mathtt{aaabbb} \cdot \mathtt{aabb}, \mathtt{aaabbb} \cdot \mathtt{aaabbb}, \dots\} \\ & \vdots \end{aligned}$$
  - Mit anderen Worten
$$L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \wedge n_2 \in \mathbb{N}_+\} .$$
  - Beachte: die Exponenten n_1 „vorne“ und n_2 „hinten“ heißen verschieden.

- Sei

$$L = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}_+\}\ ,$$

also sozusagen (immer diese Pünktchen ...)

$$L = \{\mathtt{ab}, \mathtt{aabb}, \mathtt{aaabbb}, \mathtt{aaaabbbb}, \dots\}\ .$$

- Was ist $L^2 = L \cdot L$?
  -
$$\begin{aligned} L^2 = &\ \{\mathtt{ab} \cdot \mathtt{ab}, \mathtt{ab} \cdot \mathtt{aabb}, \mathtt{ab} \cdot \mathtt{aaabbb}, \dots\} \\ &\cup \{\mathtt{aabb} \cdot \mathtt{ab}, \mathtt{aabb} \cdot \mathtt{aabb}, \mathtt{aabb} \cdot \mathtt{aaabbb}, \dots\} \\ &\cup \{\mathtt{aaabbb} \cdot \mathtt{ab}, \mathtt{aaabbb} \cdot \mathtt{aabb}, \mathtt{aaabbb} \cdot \mathtt{aaabbb}, \dots\} \\ &\ \ \vdots \end{aligned}$$
  - Mit anderen Worten

$$L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \land n_2 \in \mathbb{N}_+\}\ .$$

  - Beachte: die Exponenten n_1 „vorne“ und n_2 „hinten“ heißen verschieden.

Potenzen mehrfach definiert

- für Alphabet A und für $i \in \mathbb{N}_0$ hatten wir schon Potenzen A^i definiert.
- A^n: *Menge aller Wörter der Länge n* über dem Alphabet A
- Jedes Alphabet A kann man als formale Sprache L_A auffassen (enthält alle Wörter der Länge 1)
- Man mache sich klar:
  A^i ist („im Wesentlichen") das Gleiche wie L_A^i.
  - $A^0 = \{\varepsilon\} = L_A^0$
  - $A^1 = A = L_A = L_A^1$
  - $A^2 = A \cdot A^1 = L_A \cdot L_A^1 = L_A \cdot L_A = L_A^2$
  - ...
  - wonach sieht das aus?

Überblick

Konkatenationsabschluss von L

- bei Alphabeten schon gesehen: $A^* = \bigcup_{i=0}^{\infty} A^i$.
- der Konkatenationsabschluss L^* von L ist

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

- der ε-freie Konkatenationsabschluss L^+ von L ist

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$

- Man sieht:

$$L^* = L^0 \cup L^+ \; .$$

Einfaches Beispiel zum Konkatenationsabschluss

- $L = \{ab, c\}$
- $L^0 = \{\varepsilon\}$
- $L^1 = \{ab, c\}$
- $L^2 = \{abab, abc, cab, cc\}$
- $L^3 = \{ababab, ababc, abcab, abcc, cabab, cabc, ccab, ccc\}$
- $\ldots$
- Der Konkatenationsabschluss von L ist $L^* = \bigcup_{i=0}^{\infty} L^i$
- $L^* = L^0 \cup L^1 \cup L^2 \cup \cdots = \{\varepsilon, ab, c, abab, abc, cab, cc, \ldots\}$
- Der ε-freie Konkatenationsabschluss von L ist $L^+ = \bigcup_{i=1}^{\infty} L^i$
- $L^+ = L^1 \cup L^2 \cup \cdots = \{ab, c, abab, abc, cab, cc, \ldots\}$
- Zusammenhang: $L^* = L^0 \cup L^+$

Konkatenationsabschluss von L

- *Abgeschlossenheit*: Ist f eine n-stellige innere Verknüpfung (hier binäre/zweistellige Operation „Konkatenation") auf einer Menge A, dann heißt das: f ist eine Funktion $A^n \to A$. Gilt nun $\emptyset \neq M \subseteq A$, dann heißt M abgeschlossen bezüglich f, wenn $f(a_1, \ldots, a_n)$ in M liegt für alle $a_1, \ldots, a_n \in M$, wenn also f eingeschränkt auf den Definitionsbereich M^n auch eine n-stellige innere Verknüpfung (hier Konkatenation) auf M ist.

Weitere Beispiele für Konkatenationsabschluss

- Es sei wieder $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}_+\}$.
- haben schon gesehen:

$$L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \wedge n_2 \in \mathbb{N}_+\} .$$

- analog

$$L^3 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2}\mathtt{a}^{n_3}\mathtt{b}^{n_3} \mid n_1 \in \mathbb{N}_+ \wedge n_2 \in \mathbb{N}_+ \wedge n_3 \in \mathbb{N}_+\} .$$

- wir erlauben uns Pünktchen ...:

$$L^i = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1} \cdots \mathtt{a}^{n_i}\mathtt{b}^{n_i} \mid n_1, \dots, n_i \in \mathbb{N}_+\} .$$

- Dann kann man für L^+ notieren:

$$L^+ = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1} \cdots \mathtt{a}^{n_i}\mathtt{b}^{n_i} \mid i \in \mathbb{N}_+ \wedge n_1, \dots, n_i \in \mathbb{N}_+\} .$$

- Sie merken (hoffentlich): L^+ und L^* sind *präziser und kürzer* hinzuschreiben als viele Pünktchen.

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\texttt{a}, \texttt{b}\}$
- $L = \{\texttt{a}\}^* \cup \{\texttt{b}\}^* = \{\varepsilon, \texttt{a}, \texttt{aa}, \texttt{aaa}, \dots, \texttt{b}, \texttt{bb}, \texttt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\texttt{aa} \cdot \varepsilon \cdot \texttt{aaaa} \cdot \texttt{b} \cdot \texttt{aaaaa}$
  - z. B. $\texttt{aa} \cdot \texttt{bbbbb} \cdot \texttt{aaa} \cdot \texttt{b} \cdot \texttt{aaaaa} \cdot \texttt{bbb} \cdot \texttt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \ldots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \ldots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \ldots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus a oder nur aus b bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Konkatenationsabschluss: Beispiel

- $A = \{\mathtt{a}, \mathtt{b}\}$
- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält L^*?
  - alle Wörter, die man erhält, wenn man
    - eine beliebige endliche Zahl k
    - von Wörtern $w_1, \dots, w_k$ aus L
    - konkateniert zu $w_1 \cdots w_k$
  - z. B. $\mathtt{aa} \cdot \varepsilon \cdot \mathtt{aaaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa}$
  - z. B. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Zwei Warnungen

- Die Bezeichnung ε-freier Konkatenationsabschluss für L^+ ist irreführend.
  - Wie steht es um das leere Wort bei L^+ und L^*?
  - Klar ist:

$$\varepsilon \in L^0 \subseteq L^*$$

  - Aber: $L = L^1 \subseteq L^+$,
    wenn also $\varepsilon \in L$, dann auch $\varepsilon \in L^+$.

- Beachte

$$\{\}^* = \{\varepsilon\}$$

Zwei Warnungen

- Die Bezeichnung ε-freier Konkatenationsabschluss für L^+ ist irreführend.
  - Wie steht es um das leere Wort bei L^+ und L^*?
  - Klar ist:
    $$\varepsilon \in L^0 \subseteq L^*$$
  - Aber: $L = L^1 \subseteq L^+$,
    wenn also $\varepsilon \in L$, dann auch $\varepsilon \in L^+$.

- Beachte
  $$\{\}^* = \{\varepsilon\}$$

Was ist wichtig

Das sollten Sie mitnehmen:

- was formale Sprachen sind,
- wie ihr Produkt definiert ist und
- wie Konkatenationsabschluss und
  ε-freier Konkatenationsabschluss definiert sind.

Das sollten Sie üben:

- Erkennen von Strukturen der Form L^*, L^+, $L_1 L_2$
- Lesen von Ausdrücken der Form $(L_1^+ L_2)^*$ usw.
- „Rechnen" mit formalen Sprachen

Grundbegriffe der Informatik

Einheit 6: Dokumente

Prof. Dr. Tanja Schultz

Karlsruher Institut für Technologie, Fakultät für Informatik

Wintersemester 2012/2013

Dokumente

- LaTeX
- XHTML
- Eine Grenze unserer bisherigen Vorgehensweise

Dokumente

- Im alltäglichen Leben gibt es vielerlei Inschriften:
  Briefe, Kochrezepte, Zeitungsartikel, Vorlesungsskripte, Seiten im WWW, Emails, usw.
- oft drei verschiedene Aspekte unterscheidbar,
  die für den Leser eine Rolle spielen:
  - den Inhalt des Textes,
  - seine Struktur und
  - sein Erscheinungsbild, die (äußere) Form.

Inhalt, Struktur und Form

- den Inhalt des Textes,
- seine Struktur und
- sein Erscheinungsbild, die (äußere) Form.

- den Inhalt des Textes,
- seine Struktur und
- sein Erscheinungsbild, die (äußere) Form.

andere Form:

- den INHALT des Textes,
- seine STRUKTUR und
- sein ERSCHEINUNGSBILD, die (äußere) FORM.

Inhalt, Struktur und Form

- den Inhalt des Textes,
- seine Struktur und
- sein Erscheinungsbild, die (äußere) Form.

andere Struktur:

> *[...] den Inhalt des Textes, seine Struktur und sein Erscheinungsbild, die (äußere) Form.*

Inhalt, Struktur und Form

- den Inhalt des Textes,
- seine Struktur und
- sein Erscheinungsbild, die (äußere) Form.

anderer Inhalt:

- Balaenoptera musculus (Blauwal),
- Mesoplodon carlhubbsi (Hubbs-Schnabelwal) und
- Physeter macrocephalus (Pottwal).

Wozu Inhalt, Struktur und Form?

- Inhalt üblicherweise für Autoren und Leser im Vordergrund
  - (Ausnahmen?)
- Struktur und Form
  - sollen den Leser beim Verstehen des Inhalts unterstützen.

- Dokumente: Texte mit Inhalt, Struktur und Form
  - z. B. Programme

- ein Rat für Ihr weiteres Studium:
  - viele Dokumente (PSE, Seminar, Bachelor-Arbeit, ...)
  - Finden geeigneter Struktur hilft auch eigenem Verständnis
  - *deswegen: früh damit beginnen, etwas aufzuschreiben* (weil man dann über Struktur nachdenken muss)

Wozu Inhalt, Struktur und Form?

- Inhalt üblicherweise für Autoren und Leser im Vordergrund
  - (Ausnahmen?)
- Struktur und Form
  - sollen den Leser beim Verstehen des Inhalts unterstützen.

- Dokumente: Texte mit Inhalt, Struktur und Form
  - z. B. Programme

- ein Rat für Ihr weiteres Studium:
  - viele Dokumente (PSE, Seminar, Bachelor-Arbeit, ...)
  - Finden geeigneter Struktur hilft auch eigenem Verständnis
  - *deswegen: früh damit beginnen, etwas aufzuschreiben*
    (weil man dann über Struktur nachdenken muss)

Wozu Inhalt, Struktur und Form?

- Inhalt üblicherweise für Autoren und Leser im Vordergrund
  - (Ausnahmen?)
- Struktur und Form
  - sollen den Leser beim Verstehen des Inhalts unterstützen.

- Dokumente: Texte mit Inhalt, Struktur und Form
  - z. B. Programme

- ein Rat für Ihr weiteres Studium:
  - viele Dokumente (PSE, Seminar, Bachelor-Arbeit, ...)
  - Finden geeigneter Struktur hilft auch eigenem Verständnis
  - *deswegen: früh damit beginnen, etwas aufzuschreiben*
    (weil man dann über Struktur nachdenken muss)

- Struktur von Dokumenten:
- auch da spielt syntaktische Korrektheit eine Rolle,
- zumindest wenn Rechner im Spiel sind.
- Beispiele: Auszeichnungssprachen (engl. *markup language*)
  - Listen in LaTeX
    - und ein klein bisschen Allgemeines zu LaTeX
  - Tabellen in XHTML
  - Hypertext Markup Language (HTML; dt. Hypertext-Auszeichnungssprache) ist eine textbasierte Auszeichnungssprache zur Strukturierung von Inhalten wie Texten, Bildern und Hyperlinks in Dokumenten.
  - XHTML eXtensible HTML – erweiterbare HTML

Überblick

Dokumente

- LaTeX
- XHTML
- Eine Grenze unserer bisherigen Vorgehensweise

- basiert auf TeX von Donald Knuth
  http://www-cs-faculty.stanford.edu/~knuth/
- ausgesprochen „Tech“ (bzw. „Latech“)
- Textsatz-Programm
- in der Informatik sehr häufig verwendet wegen
  - hervorragendem automatischen Satz mathematischer Formeln
  - aus
    `\[ 2 - \sum_{i=0}^{k} i 2^{-i} = (k+2) 2^{-k} \]`
  - wird

$$2 - \sum_{i=0}^{k} i 2^{-i} = (k+2) 2^{-k}$$

- Vorlesungsskript und Folien sind auch mit LaTeX gemacht
  - Diese Folie beginnt so:

```
\begin{frame}[fragile]
  \frametitle{\LaTeX}

  \begin{itemize}
  \item basiert auf \TeX{} von Donald Knuth
```

- Im Skript steht zum Beispiel
  `\section{Struktur von Dokumenten}`
- woraus LaTeX die Zeile

  7.2 STRUKTUR VON DOKUMENTEN

  auf Seite 50 des Skriptes gemacht hat
- also
  - automatisch die passende Abschnittsnummer eingefügt
  - alles wurde in Großbuchstaben gesetzt
- Beachte: z. B. Schriftauswahl ist *nicht* in der Eingabe mit vermerkt.
- Diese Festlegung findet sich an anderer Stelle, und zwar an *einer* Stelle, an der das typografische Aussehen *aller* Abschnittsüberschriften (einheitlich) festgelegt ist.

Grobstruktur von LaTeX-Dokumenten

```
\documentclass[11pt]{report}
% so schreibt man Kommentare
% dieser Teil heißt Präambel des Dokumentes
% Unterstützung für Deutsch,
% z.B. richtige automatische Trennung
  \usepackage[german]{babel}
  % Angabe des Zeichensatzes, in dem der Text ist
  \usepackage[latin1]{inputenc}
  % für das Einbinden von Grafiken
  \usepackage{graphicx}
\begin{document}
  % und hier kommt der eigentliche Text .....
\end{document}
```

Listen mit LaTeX

- Eine Liste einfacher Punkte sieht in LaTeX so aus:

Eingabe	Ausgabe
`\begin{itemize}`	
`\item Inhalt`	• Inhalt
`\item Struktur`	• Struktur
`\item Form`	• Form
`\end{itemize}`	

- der dicke Punkt als Markierung ist *nicht* an der Stelle der Liste festgelegt.
- Trennung der Spezifikation von Struktur und der Spezifikation von Form
- Wenn man z. B. „–“ statt „•“ möchte, dann muss man an *einer* Stelle (in der Präambel) die Definition `\item` ändern.

Formale Sprachen

- Gesucht: die formale Sprache L_{itemize} aller legalen Texte für Listen in LaTeX
- Gegeben: die formale Sprache L_{item} aller Texte, die hinter einem Aufzählungspunkt vorkommen dürfen.
  (tun wir mal so ...)
- Dann
  $L_{\text{itemize}} = \{\texttt{\textbackslash begin\{itemize\}}\} \left(\{\texttt{\textbackslash item}\} L_{\text{item}}\right)^{+} \{\texttt{\textbackslash end\{itemize\}}\}$
- Problem: in Aufzählungspunkt wieder eine Liste erlaubt; also

$$L_{\text{item}} = \ldots L_{\text{itemize}} \ldots$$
$$L_{\text{itemize}} = \ldots L_{\text{item}} \ldots$$

Dokumente

- LaTeX
- XHTML
- Eine Grenze unserer bisherigen Vorgehensweise

XHTML

- HTML: Auszeichnungssprache, die man benutzt, wenn man eine WWW-Seite (be)schreibt.
- XHTML: sozusagen im wesentlichen eine noch striktere Variante von HTML
- Für beide formaler als für LaTeX festgelegt, wie syntaktisch korrekte solche Seiten aussehen.
- Das geschieht in einer sogenannten document type definition, kurz *DTD*.

Auszug aus der DTD für Tabellen in XHTML

```
<!ELEMENT table (caption?, thead?, tfoot?, (tbody+|tr+))>
<!ELEMENT caption  %Inline;>
<!ELEMENT thead    (tr)+>
<!ELEMENT tfoot    (tr)+>
<!ELEMENT tbody    (tr)+>
<!ELEMENT tr       (th|td)+>
<!ELEMENT th       %Flow;>
<!ELEMENT td       %Flow;>
```

Interpretation der DTD für Tabellen in XHTML

```
<!ELEMENT table (caption?, thead?, tfoot?, (tbody+|tr+))>
<!ELEMENT thead    (tr)+>
<!ELEMENT tfoot    (tr)+>
<!ELEMENT tbody    (tr)+>
<!ELEMENT tr       (th|td)+>
```

- `table`, `thead`, `tr`: Namen für formale Sprachen
- Bedeutung von `+` ε-freier Konkatenationsabschluss
- Bedeutung von `,`   Produkt formaler Sprachen
- Bedeutung von `|`   Vereinigung
- Fragezeichen ist neu, aber ganz einfach:

$$L^? = L^0 \cup L^1 = \{\varepsilon\} \cup L$$

  - „optionales“ Auftreten eines Wortes aus L

- Mitteilung:
  `<!ELEMENT tbody (tr)+ >` legt fest:

$$L_{\texttt{tbody}} = \{\texttt{<tbody>}\} \cdot L_{\texttt{tr}}^{+} \cdot \{\texttt{</tbody>}\}$$

- Tabellenrumpf beginnt mit `<tbody>`, endet mit `</tbody>`, dazwischen eine beliebige positive Anzahl von Tabellenzeilen
- erste Zeile aus der DTD besagt:

$$L_{\texttt{table}} = \{\texttt{<table>}\} \cdot L_{\texttt{caption}}^{?} \cdot L_{\texttt{thead}}^{?} \cdot L_{\texttt{tfoot}}^{?} \cdot L_{\texttt{tbody}}^{+} \cdot \{\texttt{</table>}\}$$

- Tabelle ist von Zeichenketten `<table>` und `</table>` umschlossen und enthält innerhalb in dieser Reihenfolge
  - optional eine Überschrift (*caption*),
  - optional einen Tabellenkopf (*table head*),
  - optional einen Tabellenfuß (*table foot*) und
  - eine beliebige positive Anzahl von Tabellenrümpfen.

Beispiel für Tabelle in XHTML

```
<table>
  <tbody>
    <tr>  <td>1</td> <td>a</td>  </tr>
    <tr>  <td>2</td> <td>b</td>  </tr>
  </tbody>
</table>
```

Überblick

Dokumente

- LaTeX
- XHTML
- Eine Grenze unserer bisherigen Vorgehensweise

Eine Grenze unserer bisherigen Vorgehensweise

- eben mit Hilfe von Produkt und Konkatenationsabschluss formaler Sprachen präzise Aussagen gemacht.
- Das ging, weil
  - etwas von einer komplizierteren Art aus Bestandteilen einfacherer Art zusammengesetzt wurde.
- Aber manchmal: größere Dinge einer Art werden zusammengesetzt aus kleineren Bestandteilen *der gleichen Art*
- Beispiel: Listen, deren Aufzählungspunkte ihrerseits wieder Listen enthalten dürfen ...

Eine Grenze unserer bisherigen Vorgehensweise

- eben mit Hilfe von Produkt und Konkatenationsabschluss formaler Sprachen präzise Aussagen gemacht.
- Das ging, weil
  - etwas von einer komplizierteren Art aus Bestandteilen einfacherer Art zusammengesetzt wurde.
- Aber manchmal: größere Dinge einer Art werden zusammengesetzt aus kleineren Bestandteilen *der gleichen Art*
- Beispiel: Listen, deren Aufzählungspunkte ihrerseits wieder Listen enthalten dürfen ...

Eine Grenze unserer bisherigen Vorgehensweise (2)

- anderes typisches Beispiel: korrekte Klammerungen (wie bei arithmetischen Ausdrücken)
- Bei einer syntaktisch korrekten Klammerung gibt es zu jeder *Klammer auf* „weiter hinten" die „zugehörige" *Klammer zu*.
- insbesondere legal:
  - mehrere korrekte Klammerungen hintereinander
  - zusätzliche Klammern um korrekte Klammerung außen herum
- schön wäre z. B. Beziehung von L_{Klammer} mit L_{Klammer}^* und mit $\{\,(\,\} \cdot L_{\text{Klammer}} \cdot \{\,)\,\}$
- aber wie? Gleichung?
  - lösbar?
  - wenn ja: eindeutig?
  - ein Thema des nächsten Kapitels ...

Zusammenfassung

- Dokumente haben Inhalt, Struktur und Form.
- formale Sprachen helfen bei der Definition legaler Strukturen.