Grundbegriffe der Informatik

Kapitel 7: formale Sprachen

Mattias Ulbrich
(basierend auf Folien von Thomas Worsch)

KIT · Institut für Theoretische Informatik

Wintersemester 2023/2024

Erinnerung: formale Sprache

- *formale Sprache über einem Alphabet A*
  eine Teilmenge $L \subseteq A^*$

Wo sind wir?

Produkt formaler Sprachen

Konkatenationsabschluss formaler Sprachen

Produkt oder Konkatenation formaler Sprachen

- „passend“ zur Definition bei Wörtern
- Produkt der Sprachen L_1 und L_2:

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

- wegen der Assoziativität der Konkatenation von Wörtern auch

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

- den Punkt lässt man auch gerne wieder weg

Produkte formaler Sprachen — Beispiele (1)

- wenn $L_1 = \{\mathtt{a}\}$ und $L_2 = \{\mathtt{b}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}\} \cdot \{\mathtt{b}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}\} \\ &= \{\mathtt{ab}\} \end{aligned}$$

Produkte formaler Sprachen — Beispiele (1)

- wenn $L_1 = \{\mathtt{a}\}$ und $L_2 = \{\mathtt{b}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}\} \cdot \{\mathtt{b}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}\} \\ &= \{\mathtt{ab}\} \end{aligned}$$
- wenn $L_1 = \{\mathtt{a}\}$ und $L_2 = \{\mathtt{b}, \mathtt{ab}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}\} \cdot \{\mathtt{b}, \mathtt{ab}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}, \mathtt{a} \cdot \mathtt{ab}\} \\ &= \{\mathtt{ab}, \mathtt{aab}\} \end{aligned}$$

Produkte formaler Sprachen — Beispiele (1)

- wenn $L_1 = \{\mathtt{a}\}$ und $L_2 = \{\mathtt{b}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}\} \cdot \{\mathtt{b}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}\} \\ &= \{\mathtt{ab}\} \end{aligned}$$
- wenn $L_1 = \{\mathtt{a}\}$ und $L_2 = \{\mathtt{b}, \mathtt{ab}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}\} \cdot \{\mathtt{b}, \mathtt{ab}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}, \mathtt{a} \cdot \mathtt{ab}\} \\ &= \{\mathtt{ab}, \mathtt{aab}\} \end{aligned}$$
- wenn $L_1 = \{\mathtt{a}, \mathtt{aa}\}$ und $L_2 = \{\mathtt{b}, \mathtt{ab}\}$
  dann
  $$\begin{aligned} L_1 \cdot L_2 &= \{\mathtt{a}, \mathtt{aa}\} \cdot \{\mathtt{b}, \mathtt{ab}\} \\ &= \{\mathtt{a} \cdot \mathtt{b}, \mathtt{a} \cdot \mathtt{ab}, \mathtt{aa} \cdot \mathtt{b}, \mathtt{aa} \cdot \mathtt{ab}\} \\ &= \{\mathtt{ab}, \mathtt{aab}, \mathtt{aaab}\} \end{aligned}$$

Produkte formaler Sprachen — Beispiele (2)

- betrachte ASCII-Alphabet und Sprachen
  $S = \{\texttt{int}, \texttt{double}, \texttt{char}\}$,
  $B = \{\texttt{a}, \ldots, \texttt{z}\}$ und
  $Z = \{\texttt{0}, \ldots, \texttt{9}\}$
- $S \cdot \{\texttt{␣}\} \cdot B \cdot (B \cup Z \cup \{\varepsilon\}) \cdot \{\texttt{;}\}$

Produkte formaler Sprachen — Beispiele (2)

- betrachte ASCII-Alphabet und Sprachen
  $S = \{\texttt{int}, \texttt{double}, \texttt{char}\}$,
  $B = \{\texttt{a}, \ldots, \texttt{z}\}$ und
  $Z = \{\texttt{0}, \ldots, \texttt{9}\}$
- $S \cdot \{\texttt{␣}\} \cdot B \cdot (B \cup Z \cup \{\varepsilon\}) \cdot \{\texttt{;}\}$
  enthält „Deklarationen" wie z. B.
  - `int␣x2;`
  - `double␣w;`

Produkte formaler Sprachen — Beispiele (2)

- betrachte ASCII-Alphabet und Sprachen
  $S = \{\texttt{int}, \texttt{double}, \texttt{char}\}$,
  $B = \{\texttt{a}, \ldots, \texttt{z}\}$ und
  $Z = \{\texttt{0}, \ldots, \texttt{9}\}$
- $S \cdot \{\texttt{␣}\} \cdot B \cdot (B \cup Z \cup \{\varepsilon\}) \cdot \{\texttt{;}\}$
  enthält „Deklarationen“ wie z. B.
  - `int␣x2;`
  - `double␣w;`

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

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

$$L \cdot \{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L \text{ und } w_2 \in \{\varepsilon\}\}$$

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

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

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

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

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

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

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

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

Neutrales Element für Produkt formaler Sprachen

Lemma

- Für jede formale Sprache L ist

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

Beweis

- einfaches Nachrechnen:

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

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

Potenzen von Sprachen — wie bei Wörtern induktiv definiert

- „Problem“: Was soll L^0 sein?

Potenzen von Sprachen — wie bei Wörtern induktiv definiert

- „Problem": Was soll L^0 sein?
- Definiere:

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

Potenzen von Sprachen —
wie bei Wörtern induktiv definiert

Karlsruher Institut für Technologie

- „Problem“: Was soll L^0 sein?
- Definiere:

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

- 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

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

Beispiele für Potenzen von Sprachen (2)

- 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 $L^2 = L \cdot L$?
  - $$\begin{aligned} L^2 = \ & \{\mathtt{ab} \cdot \mathtt{ab}, \mathtt{ab} \cdot \mathtt{aabb}, \mathtt{ab} \cdot \mathtt{aaabbb}, \dots\} \\ & \cup \{\mathtt{aabb} \cdot \mathtt{ab}, \mathtt{aabb} \cdot \mathtt{aabb}, \mathtt{aabb} \cdot \mathtt{aaabbb}, \dots\} \\ & \cup \{\mathtt{aaabbb} \cdot \mathtt{ab}, \mathtt{aaabbb} \cdot \mathtt{aabb}, \mathtt{aaabbb} \cdot \mathtt{aaabbb}, \dots\} \\ & \cup \cdots \end{aligned}$$

Beispiele für Potenzen von Sprachen (2)

- 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 $L^2 = L \cdot L$?
  - $$\begin{aligned} L^2 = \quad & \{\mathtt{ab}\cdot\mathtt{ab}, \mathtt{ab}\cdot\mathtt{aabb}, \mathtt{ab}\cdot\mathtt{aaabbb}, \dots\} \\ \cup\ & \{\mathtt{aabb}\cdot\mathtt{ab}, \mathtt{aabb}\cdot\mathtt{aabb}, \mathtt{aabb}\cdot\mathtt{aaabbb}, \dots\} \\ \cup\ & \{\mathtt{aaabbb}\cdot\mathtt{ab}, \mathtt{aaabbb}\cdot\mathtt{aabb}, \mathtt{aaabbb}\cdot\mathtt{aaabbb}, \dots\} \\ \cup\ & \cdots \end{aligned}$$
  - also $\quad L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+\}\ .$
  - Exponenten n_1 und n_2 können verschieden sein

Potenzen mehrfach definiert

- für Alphabet A schon Potenzen A^i definiert.
- Jedes Alphabet A kann man als formale Sprache L_A auffassen
  - enthält alle Wörter der Länge 1
- Man mache sich klar:
  A^i ist «das Gleiche» wie L_A^i.

Wo sind wir?

Produkt formaler Sprachen

Konkatenationsabschluss formaler Sprachen

Konkatenationsabschluss von L — beliebig viele Faktoren aus L

Karlsruher Institut für Technologie

- bei Alphabeten: $A^* = \bigcup_{i \in \mathbb{N}_0} A^i$.
- Konkatenationsabschluss L^* von L (kleenesche Hülle, Kleene Star)

$$L^* = \bigcup_{i \in \mathbb{N}_0} L^i$$

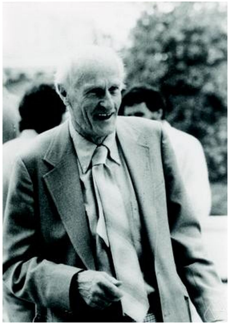
- ε-freier Konkatenationsabschluss L^+ von L (Kleene-+-Abschluss)

$$L^+ = \bigcup_{i \in \mathbb{N}_+} L^i$$

- Man sieht:

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

Stephen Cole Kleene



Bildquelle https://commons.wikimedia.org/wiki/File:Kleene.jpg

- geboren 5. Januar 1909 in Hartford, CT, USA
  gestorben am 25. Januar 1994 in Madison, WI, USA
- einer der Begründer der theoretischen Informatik
- insbesondere: formale Sprachen, kleenesche Hülle, Lambda-Kalkül, reguläre Ausdrücke

Beispiele für Konkatenationsabschluss (1)

- Es sei wieder $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}_+\}$.
- schon gesehen
  $L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+\}$
- analog
  $L^3 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2}\mathtt{a}^{n_3}\mathtt{b}^{n_3} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+ \text{ und } n_3 \in \mathbb{N}_+\}$
- mit 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
  $L^+ = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1} \cdots \mathtt{a}^{n_i}\mathtt{b}^{n_i} \mid i \in \mathbb{N}_+ \text{ und } n_1, \dots, n_i \in \mathbb{N}_+\}$
- L^+ und L^* sind *präziser und kürzer* als viele Pünktchen

Beispiele für Konkatenationsabschluss (2)

$A = \{\mathtt{a}, \mathtt{b}\}$

- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \ldots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \ldots\}$
- Welche Wörter enthält $L^* = (\{\mathtt{a}\}^* \cup \{\mathtt{b}\}^*)^*$?
  - 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. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$

Beispiele für Konkatenationsabschluss (2)

Karlsruher Institut für Technologie

$A = \{\mathtt{a}, \mathtt{b}\}$

- $L = \{\mathtt{a}\}^* \cup \{\mathtt{b}\}^* = \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \dots, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \dots\}$
- Welche Wörter enthält $L^* = (\{\mathtt{a}\}^* \cup \{\mathtt{b}\}^*)^*$?
  - 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. $\mathtt{aa} \cdot \mathtt{bbbbb} \cdot \mathtt{aaa} \cdot \mathtt{b} \cdot \mathtt{aaaaa} \cdot \mathtt{bbb} \cdot \mathtt{aaa}$
- Beobachtung
  - jedes Wort aus A^* ist die Konkatenation von Blöcken, die nur aus $\mathtt{a}$ oder nur aus $\mathtt{b}$ bestehen
  - also ist $L^* = A^*$

Beispiele für Konkatenationsabschluss (3)

- wenn
  - $S = \{\texttt{int}, \texttt{double}, \texttt{char}\}$
  - $B = \{\mathtt{a}, \ldots, \mathtt{z}\}$
  - $Z = \{\mathtt{0}, \ldots, \mathtt{9}\}$
- dann enthält

$$S \cdot \{\texttt{␣}\}^+ \cdot B \cdot (B \cup Z)^* \cdot \{\texttt{␣}\}^* \cdot \{\texttt{;}\}$$

„Deklarationen“ wie z. B.

  - `int␣␣x42␣;`
  - `double␣wurzelzwei;`

Beispiele für Konkatenationsabschluss (3)

- wenn
  - $S = \{\texttt{int}, \texttt{double}, \texttt{char}\}$
  - $B = \{\texttt{a}, \ldots, \texttt{z}\}$
  - $Z = \{\texttt{0}, \ldots, \texttt{9}\}$
- dann enthält

$$S \cdot \{\texttt{␣}\}^+ \cdot B \cdot (B \cup Z)^* \cdot \{\texttt{␣}\}^* \cdot \{\texttt{;}\}$$

  „Deklarationen" wie z. B.
  - `int␣␣x42␣;`
  - `double␣wurzelzwei;`
- aber leider auch
  - `int␣double;` usw.

Zwei Warnungen

- L^+ „ε-freier“ Konkatenationsabschluss: irreführend
  - klar ist:
    $$\varepsilon \in L^0 \subseteq L^*$$
  - aber: $L = L^1 \subseteq L^+$,
    wenn also $\varepsilon \in L$, dann auch $\varepsilon \in L^+$.

Zwei Warnungen

Karlsruher Institut für Technologie

- L^+ „ε-freier“ Konkatenationsabschluss: irreführend
  - 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_1L_2
  - Lesen von Ausdrücken der Form $(L_1^+L_2)^*$ usw.
  - „Rechnen“ mit formalen Sprachen