Automatentheorie und Formale Sprachen
Mengen, Alphabete, Wörter, formale Sprachen

Dozentin: Wiebke Petersen

29.4.2009

Mengen

Definition 1. *Eine* **Menge** *ist eine Zusammenfassung beliebiger Objekte, genannt Elemente, zu einer Gesamtheit, wobei keines der Objekte die Menge selbst sein darf. Zwei Mengen sind* **gleich***, g.d.w. sie die gleichen Elemente enthalten. Es gibt genau eine Menge, die keine Elemente enthält, die* **leere Menge** $\emptyset$.

Mengen

Definition 1. *Eine* ***Menge*** *ist eine Zusammenfassung beliebiger Objekte, genannt Elemente, zu einer Gesamtheit, wobei keines der Objekte die Menge selbst sein darf. Zwei Mengen sind* ***gleich****, g.d.w. sie die gleichen Elemente enthalten. Es gibt genau eine Menge, die keine Elemente enthält, die* ***leere Menge*** $\emptyset$.

explizite Mengendarstellung $\{a_1, a_2, \ldots, a_n\}$ ist die Menge, die genau die Elemente $a_1, a_2, \ldots, a_n$ enthält.
Beispiel: $\{2, 3, 4, 5, 6, 7\}$

Mengen

Definition 1. *Eine* **Menge** *ist eine Zusammenfassung beliebiger Objekte, genannt Elemente, zu einer Gesamtheit, wobei keines der Objekte die Menge selbst sein darf. Zwei Mengen sind* **gleich***, g.d.w. sie die gleichen Elemente enthalten. Es gibt genau eine Menge, die keine Elemente enthält, die* **leere Menge** $\emptyset$*.*

explizite Mengendarstellung $\{a_1, a_2, \ldots, a_n\}$ ist die Menge, die genau die Elemente $a_1, a_2, \ldots, a_n$ enthält.
Beispiel: $\{2, 3, 4, 5, 6, 7\}$

implizite Mengendarstellung $\{x|A\}$ ist die Menge, die genau die Objekte x enthält, auf die die Aussage A zutrifft.
Beispiel: $\{x|x \in \mathbb{N}$ und $x < 8$ und $1 < x\}$

wichtige Mengennotationen

$Z = \{x|x \text{ ist gerade }\}$, $D = \{x|x \text{ ist glatt durch 3 teilbar}\}$,
$S = \{x|x \text{ ist glatt durch 6 teilbar}\}$

- $\boldsymbol{x \in M}$: x ist ein **Element** der Menge M ($2 \in Z$, $2 \notin D$)
- $N \subset \boldsymbol{M}$: die Menge N ist eine **Teilmenge** der Menge M ($D \subset S$)
- $\boldsymbol{N \cap M}$: die **Schnittmenge** der Mengen M und N ($D \cap Z = S$)
- $\boldsymbol{N \cup M}$: die **Vereinigungsmenge** der Mengen M und N ($S \cup D = D$, $Z \cup D = \{x|x \in Z \text{ oder } x \in D\}$)
- $\boldsymbol{\mathcal{POT}(M)}$: die **Potenzmenge** der Menge M ist die Menge aller Teilmengen von M, also $\mathcal{POT}(M) = \{N|N \subseteq M\}$.
  Für endliche Mengen gilt: ist M eine n-elementige Menge, so ist $\mathcal{POT}(M)$ eine 2^n-elementige Menge.

Logische Operatoren und Quantoren

$Z = \{x|x \text{ ist gerade}\}$, $D = \{x|x \text{ ist glatt durch 3 teilbar}\}$,
$S = \{x|x \text{ ist glatt durch 6 teilbar}\}$

$\wedge$ **(und):** Wenn $x \in Z \wedge x \in D$ dann $x \in S$.

$\vee$ **(oder):** Wenn $x \in Z \vee x \in D$ dann $x \neq 5$.

$\longrightarrow$ **(impliziert):** $x \in S \rightarrow x \in Z$.

Logische Operatoren und Quantoren

$Z = \{x|x \text{ ist gerade}\}$, $D = \{x|x \text{ ist glatt durch 3 teilbar}\}$,
$S = \{x|x \text{ ist glatt durch 6 teilbar}\}$

$\wedge$ **(und):** Wenn $x \in Z \wedge x \in D$ dann $x \in S$.

$\vee$ **(oder):** Wenn $x \in Z \vee x \in D$ dann $x \neq 5$.

$\longrightarrow$ **(impliziert):** $x \in S \rightarrow x \in Z$.

$\forall$ **(Allquantor):** $\forall x \in D$: die Quersumme von x ist durch 3 teilbar.

$\exists$ **(Existenzquantor):** $\exists x \in D$: die Quersumme von x ist 9.

Alphabete

Ein **Alphabet** ist eine nicht-leere endliche Menge von *Zeichen* (Symbolen, Buchstaben).

Für Alphabete verwenden wir in der Regel $\Sigma, \Gamma, \ldots$, während $a, b, c, \ldots$ i.d.R. für die Zeichen stehen.

Beispiele:

Alphabete

Ein **Alphabet** ist eine nicht-leere endliche Menge von *Zeichen* (Symbolen, Buchstaben).

Für Alphabete verwenden wir in der Regel $\Sigma, \Gamma, \ldots$, während $a, b, c, \ldots$ i.d.R. für die Zeichen stehen.

Beispiele:

- lateinisches Alphabet $\Sigma_{lat} = \{A, B, C, \ldots, Z\}$
- das Morsealphabet $\Sigma_M = \{\cdot, -, \sqcup\}$
- das Ziffernalphabet $\Sigma_Z = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$

Wörter

Ein **Wort** ist eine endliche Folge/Kette $x_1 \dots x_n$ von Zeichen eines Alphabets $(n \geq 0)$.
Wörter bezeichnen wir in der Regel mit $u, v, w, \dots$.

Wörter

Ein **Wort** ist eine endliche Folge/Kette $x_1 \dots x_n$ von Zeichen eines Alphabets $(n \geq 0)$.
Wörter bezeichnen wir in der Regel mit $u, v, w, \dots$.

Die Menge aller Wörter über einem Alphabet Σ bezeichnen wir mit Σ^*.

Wörter

Ein **Wort** ist eine endliche Folge/Kette $x_1 \ldots x_n$ von Zeichen eines Alphabets $(n \geq 0)$.
Wörter bezeichnen wir in der Regel mit $u, v, w, \ldots$.

Die Menge aller Wörter über einem Alphabet Σ bezeichnen wir mit Σ^*.

ϵ bezeichnet das leere Wort, welches aus null Zeichen besteht.

Wörter

Ein **Wort** ist eine endliche Folge/Kette $x_1 \ldots x_n$ von Zeichen eines Alphabets $(n \geq 0)$.
Wörter bezeichnen wir in der Regel mit $u, v, w, \ldots$.

Die Menge aller Wörter über einem Alphabet Σ bezeichnen wir mit Σ^*.

ϵ bezeichnet das leere Wort, welches aus null Zeichen besteht.

$\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$ ist die Menge der nichtleeren Wörter.

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| =$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a =$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$
- $|ababbb|_b =$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$
- $|ababbb|_b = 4$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a =$ 2
- $|ababbb|_b =$ 4
- $|ababbb|_c =$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$
- $|ababbb|_b = 4$
- $|ababbb|_c = 0$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:

(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$
- $|ababbb|_b = 4$
- $|ababbb|_c = 0$
- $|\epsilon| =$

Länge von Wörtern

$|w|$ ist die Länge von w, also die Zahl der Zeichen in dem Wort w, wobei mehrfach vorkommende Zeichen mehrfach gezählt werden.

$|w|_a$ ist die Zahl des Vorkommens des Zeichens a in dem Wort w.

Die Länge eines Wortes kann induktiv definiert werden:
(1) $|\epsilon| = 0$ und
(2) $|wa| = 1 + |w|$ für $a \in \Sigma$ und $w \in \Sigma^*$.

Beispiele:

- $|ababbb| = 6$
- $|ababbb|_a = 2$
- $|ababbb|_b = 4$
- $|ababbb|_c = 0$
- $|\epsilon| = 0$

Leersymbol, leeres Wort und leere Menge

Vorsicht Verwechslungsgefahr!

Leersymbol, leeres Wort und leere Menge

Vorsicht Verwechslungsgefahr!

Das Leersymbol ␣ ist ein *Zeichen* des Alphabets, also auch ein Wort der Länge 1.

Leersymbol, leeres Wort und leere Menge

Vorsicht Verwechslungsgefahr!

Das Leersymbol $\sqcup$ ist ein *Zeichen* des Alphabets, also auch ein Wort der Länge 1.

Das leere Wort ϵ ist ein *Wort* der Länge 0.

Leersymbol, leeres Wort und leere Menge

Vorsicht Verwechslungsgefahr!

Das Leersymbol $\sqcup$ ist ein *Zeichen* des Alphabets, also auch ein Wort der Länge 1.

Das leere Wort ϵ ist ein *Wort* der Länge 0.

Die leere Menge $\emptyset$ ist eine *Menge*.

Verkettung / Konkatenation

Die Verkettung zweier Wörter $w = a_1a_2 \dots a_n$ und $v = b_1b_2 \dots b_m$ mit $n, m \geq 0$ ist

$$w \circ v = a_1 \dots a_n b_1 \dots b_m$$

Die Verkettung $\circ$ ist also eine Funktion $\circ : \Sigma^* \times \Sigma^* \to \Sigma^*$, die Paaren von Wörtern neue Wörter zuordnet.

Verkettung / Konkatenation

Die Verkettung zweier Wörter $w = a_1a_2 \ldots a_n$ und $v = b_1b_2 \ldots b_m$ mit $n, m \geq 0$ ist

$$w \circ v = a_1 \ldots a_n b_1 \ldots b_m$$

Die Verkettung $\circ$ ist also eine Funktion $\circ : \Sigma^* \times \Sigma^* \to \Sigma^*$, die Paaren von Wörtern neue Wörter zuordnet.

Beachte, daß folgende Rechenregeln gelten:

$$w \circ \epsilon = \epsilon \circ w = w \quad \text{neutrales Element}$$

$$u \circ (v \circ w) = (u \circ v) \circ w \quad \text{Assoziativität}$$

Die Struktur $(\Sigma^*, \circ)$ bildet eine Halbgruppe mit neutralem Element (Monoid).
Vereinfachend schreibt man statt $u \circ v$ auch uv.

Potenzen von Wörtern

Die n-te Potenz eines Wortes w ist die n-fache Verkettung von w mit sich selbst.

Induktive Definition:

$$w^0 = \epsilon, \;\; w^{n+1} = w^n \circ w$$

Potenzen von Wörtern

Die n-te Potenz eines Wortes w ist die n-fache Verkettung von w mit sich selbst.

Induktive Definition:

$$w^0 = \epsilon, \;\; w^{n+1} = w^n \circ w$$

Beispiel: $w^3 = w \circ w \circ w$

Die Iteration (Kleene-Stern) von w ist

$$w^* := \bigcup_{n \geq 0} \{w^n\}$$

Für jedes beliebige Wort w gilt: $\epsilon \in w^*$

formale Sprachen

Eine **formale Sprache** L ist eine Menge von Wörtern über einem Alphabet Σ, also $L \subseteq \Sigma^*$.

formale Sprachen

Eine **formale Sprache** L ist eine Menge von Wörtern über einem Alphabet Σ, also $L \subseteq \Sigma^*$.

Beispiele:

- Sprache L_{rom} der gültigen römischen Zahldarstellungen über dem Alphabet $\Sigma_{rom} = \{\mathbf{I}, \mathbf{V}, \mathbf{X}, \mathbf{L}, \mathbf{C}, \mathbf{D}, \mathbf{M}\}$.

formale Sprachen

Eine **formale Sprache** L ist eine Menge von Wörtern über einem Alphabet Σ, also $L \subseteq \Sigma^*$.

Beispiele:

- Sprache L_{rom} der gültigen römischen Zahldarstellungen über dem Alphabet $\Sigma_{rom} = \{\mathbf{I}, \mathbf{V}, \mathbf{X}, \mathbf{L}, \mathbf{C}, \mathbf{D}, \mathbf{M}\}$.
- Sprache L_{Mors} der Buchstaben des lateinischen Alphabets dargestellt im Morsecode. $L_{Mors} = \{\cdot -, - \cdot \cdot\cdot, \ldots, - - \cdot\cdot\}$

formale Sprachen

Eine **formale Sprache** L ist eine Menge von Wörtern über einem Alphabet Σ, also $L \subseteq \Sigma^*$.

Beispiele:

- Sprache L_{rom} der gültigen römischen Zahldarstellungen über dem Alphabet $\Sigma_{rom} = \{\mathbf{I}, \mathbf{V}, \mathbf{X}, \mathbf{L}, \mathbf{C}, \mathbf{D}, \mathbf{M}\}$.
- Sprache L_{Mors} der Buchstaben des lateinischen Alphabets dargestellt im Morsecode. $L_{Mors} = \{\cdot -, - \cdot \cdot\cdot, \ldots, - - \cdot\cdot\}$
- Sprache L_{arith} der vollständig geklammerten arithmetischen Ausdrücke über dem Alphabet $\Sigma_Z \cup \{(,), +, -, \cdot, :\}$.

Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet Σ, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über Σ:

$$K \cup L,\ K \cap L,\ K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L =$
  $L \circ K =$
- $K \circ \emptyset =$
- $K \circ \{\epsilon\} =$

Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet Σ, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über Σ:

$$K \cup L, \; K \cap L, \; K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L = \{abbbbb, abbab, abbb, aab\}$ und
  $L \circ K =$
- $K \circ \emptyset =$
- $K \circ \{\epsilon\} =$

Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet Σ, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über Σ:

$$K \cup L,\ K \cap L,\ K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L = \{abbbbb, abbab, abbb, aab\}$ und
  $L \circ K = \{bbbabb, bbba, ababb, aba\}$
- $K \circ \emptyset =$
- $K \circ \{\epsilon\} =$

Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet Σ, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über Σ:

$$K \cup L,\ K \cap L,\ K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L = \{abbbbb, abbab, abbb, aab\}$ und
  $L \circ K = \{bbbabb, bbba, ababb, aba\}$
- $K \circ \emptyset = \emptyset$
- $K \circ \{\epsilon\} =$

Operationen auf Sprachen

Seien $L \subseteq \Sigma^*$ und $K \subseteq \Sigma^*$ zwei Sprachen über dem Alphabet Σ, dann entstehen durch die Verknüpfung mit Mengenoperatoren neue Sprachen über Σ:

$$K \cup L,\ K \cap L,\ K \setminus L$$

Die Verkettung von Wörtern kann ausgedehnt werden auf die Verkettung von Sprachen:

$$K \circ L := \{v \circ w \in \Sigma^* | v \in K, w \in L\}$$

Beispiel: Sei $K = \{abb, a\}$ und $L = \{bbb, ab\}$

- $K \circ L = \{abbbbb, abbab, abbb, aab\}$ und
  $L \circ K = \{bbbabb, bbba, ababb, aba\}$
- $K \circ \emptyset = \emptyset$
- $K \circ \{\epsilon\} = K$

Potenzen von Sprachen, Iteration, Kleene-Stern

Die n-te Potenz einer Sprache L ist die n-fache Verkettung von L mit sich selbst:

$$L^n = \underbrace{L \circ L \circ L \ldots \circ L}_{n\text{-mal}}$$

Induktive Definition:

$$L^0 = \{\epsilon\}, \;\; L^{n+1} = L^n \circ L$$

Potenzen von Sprachen, Iteration, Kleene-Stern

Die n-te Potenz einer Sprache L ist die n-fache Verkettung von L mit sich selbst:

$$L^n = \underbrace{L \circ L \circ L \ldots \circ L}_{n\text{-mal}}$$

Induktive Definition:

$$L^0 = \{\epsilon\}, \;\; L^{n+1} = L^n \circ L$$

Die Iteration (Kleene-Stern) von L ist

$$L^* := \bigcup_{n \geq 0} L^n$$

Für jede beliebige Sprache L gilt: $\{\epsilon\} \in L^*$

Aufgaben

Sei $K = \{aa, aaaa, ab\}$, $L = \{bb, aa\}$

1. Geben sie die Sprachen $K \circ L$, $L \circ K$, $\{\epsilon\} \circ L$, $\{\epsilon\} \circ \emptyset$ und $K \circ \emptyset$ an.

2. Geben sie die Sprache K^3 an.

3. Geben sie die Sprache $K \setminus L$ an.

4. Geben sie eine implizite Mengendarstellung der Sprache $K \circ L$ an.

5. Wie unterscheiden sich die Sprachen L^* und L^+?