Übungsstunde 1

Dienstag, 26. September 2023 11:28

Motivation Warum sind Alphabete, Wörter & Sprachen für die Informatik interessant?

In der algorithmischen Datenverarbeitung können wir Programme als Texte über dem Alphabet der Rechnertastatur verstehen. Jedes Programm realisiert also eine Transformation von Eingabetexten in Ausgabetexte. Wir starten damit den Formalismus für den Umgang mit Texten als Informationsträger einzuführen

$\mathbb{N} = \{0, 1, 2, \ldots\}$ in diesem Kurs!

Theorie

Definition 2.1. *Eine endliche nichtleere Menge Σ heißt* **Alphabet**. *Die Elemente eines Alphabets werden* **Buchstaben** (**Zeichen, Symbole**) *genannt.* **Note:** $\lambda \notin \Sigma$

Wichtig: Es dürfen beliebige, aber nur endlich viele Symbole für das Alphabet verwendet werden

Bsp.: $\Sigma_{bool} = \{0, 1\}$, $\Sigma_{lat} = \{a, \ldots, z\}$, $\Sigma_{tastatur} = \Sigma_{lat} \cup \{A, \ldots, Z, \sqcup, >, <, \ldots\}$

Definition 2.2. *Sei Σ ein Alphabet. Ein* **Wort** *über Σ ist eine endliche (eventuell leere) Folge von Buchstaben aus Σ. Das* **leere Wort** $\boldsymbol{\lambda}$ *ist die leere Buchstabenfolge. (Manchmal benutzt man ε statt λ.)*

Die **Länge** $|\boldsymbol{w}|$ *eines Wortes w ist die Länge des Wortes als Folge, d. h. die Anzahl der Vorkommen von Buchstaben in w.*

$\boldsymbol{\Sigma^*}$ *ist die Menge aller Wörter über Σ,* $\boldsymbol{\Sigma^+} = \Sigma^* - \{\lambda\}$ (nicht-leere Wörter)

Def.: Seien $x \in \Sigma^*$ und $a \in \Sigma$. Dann ist $|x|_a$ die Anzahl der Vorkommen von a in x

⚠ **Metavariable:** a ist ein beliebiger Buchstabe in Σ und <u>nicht</u> der Buchstabe "a" welcher in Σ vorkommen könnte

⚠ endlich

Bem: • Wir schreiben Wörter ohne Komma, i.e. eine Folge $x_1, x_2, \ldots, x_n$ schreiben wir immer als $x_1 x_2 \ldots x_n$

• $|\lambda| = 0$, $|\sqcup| = 1$ in $\Sigma_{tastatur}$

Bsp. Wörter: a, ab, $aabb$, $abab$

Definition 2.3. *Die* **Verkettung** (**Konkatenation**) *für ein Alphabet Σ ist eine Abbildung* $\mathrm{Kon}\colon \Sigma^* \times \Sigma^* \to \Sigma^*$, *so dass*

$$\mathrm{Kon}(x, y) = x \cdot y = xy$$

für alle $x, y \in \Sigma^*$.

Wichtig: Konkanetation ist assoziativ und $|xy| = |x| + |y|$

Bsp.: $\Sigma = \{a,b,c\}$. Sei $x = abba$, $y = cbcbc$, $z = aaac$

$Kon(x, Kon(y,z)) = Kon(x, yz) = xyz = abbacbcbcaaac$

$|xy| = |abbacbcbc| = 9$, $|x| + |y| = 4 + 5 = 9$ ✓

Definition 2.6. *Seien $v, w \in \Sigma^*$ für ein Alphabet Σ.*

- *v heißt ein **Teilwort** von w $\iff$ $\exists x, y \in \Sigma^* : w = xvy$.*
- *v heißt ein **Präfix** von w $\iff$ $\exists y \in \Sigma^* : w = vy$.*
- *v heißt ein **Suffix** von w $\iff$ $\exists x \in \Sigma^* : w = xv$.*
- *$v \neq \lambda$ heißt ein **echtes** Teilwort (Präfix, Suffix) von w genau dann, wenn $v \neq w$ und v ein Teilwort (Präfix, Suffix) von w ist.*

Bsp.: $\Sigma = \{a,b,c\}$. Sei $x = abba$, $y = cbcbc$

- bc ist ein echter Suffix von y
- $abba$ ist kein echtes Teilwort von x, aber Präfix, Suffix & Teilwort von x

Definition 2.9. *Eine **Sprache** L über einem Alphabet Σ ist eine Teilmenge von Σ^*. Das Komplement $\boldsymbol{L^C}$ der Sprache L bezüglich Σ ist die Sprache $\Sigma^* - L$.*

*$\boldsymbol{L_\emptyset} = \emptyset$ ist die **leere Sprache**.*

$\boldsymbol{L_\lambda} = \{\lambda\}$ ist die einelementige Sprache, die nur aus dem leeren Wort besteht.

Sind L_1 und L_2 Sprachen über Σ, so ist

$$\boldsymbol{L_1 \cdot L_2 = L_1 L_2} = \{vw \mid v \in L_1 \text{ und } w \in L_2\}$$

*die **Konkatenation** von L_1 und L_2. Ist L eine Sprache über Σ, so definieren wir*

$$\boldsymbol{L^0} := L_\lambda \text{ und } \boldsymbol{L^{i+1}} = L^i \cdot L \text{ für alle } i \in \mathbb{N},$$

$$\boldsymbol{L^*} = \bigcup_{i \in \mathbb{N}} L^i \text{ und } \boldsymbol{L^+} = \bigcup_{i \in \mathbb{N} - \{0\}} L^i = L \cdot L^*.$$

*L^* nennt man den **Kleene'schen Stern** von L.*

Aufgabe: Betrachte das Alphabet $\Sigma = \{a, b, c\}$ und $n \in \mathbb{N}$. Bestimme die Anzahl der Wörter aus Σ^n (= Wörter der Länge n), die a als Teilwort enthalten.

Beweis: Oft ist es einfacher zunächst das Gegenteil zu zeigen. Wir zählen also wie viele Wörter aus Σ^n den Buchstaben a nicht enthalten. Für jede der n Positionen kann also entweder der Buchstabe b oder c gewählt werden. Aber dies sind genau alle möglichen Wörter der Länge n aus dem Alphabet $\{b, c\}$.

Da $|\{b,c\}^n| = 2^n$ folgern wir, dass es genau 2^n Wörter in Σ^n gibt, welche den Buchstaben a nicht enthalten. Aus $|\Sigma^n| = 3^n$ schliessen wir, dass es $3^n - 2^n$ Wörter in Σ^n gibt, welche das Teilwort a enthalten. $\square$

Lemma 2.2. *Seien L_1, L_2, L_3 Sprachen über einem Alphabet Σ. Dann gilt*

$$L_1(L_2 \cap L_3) \subseteq L_1L_2 \cap L_1L_3.$$

Wieso nicht "="? $\Sigma = \{0,1\}$, $L_1 = \{\lambda, 1\}$, $L_2 = \{0\}$, $L_3 = \{10\}$

Dann: $L_1(L_2 \cap L_3) = \emptyset \neq \{10\} = L_1L_2 \cap L_1L_3$

Aufgabe: Beweise oder widerlege

a) Falls für eine Sprache $L^2 = L$ gilt, dann gilt auch $L^3 = L$

b) Es gibt eine Sprache L, die die Bedingung $L^2 = L$ erfüllt

c) Seien L_1, L_2, L_3 Sprachen über $\Sigma = \{0\}$. Gilt $L_1 \cdot (L_2 \cap L_3) = L_1L_2 \cap L_1L_3$?

d) $L_2 \cdot (L_1 - L_3) = \underset{L_2L_1 \cap (L_2L_3)^c}{L_2L_1 - L_2L_3}$, falls $L_1, L_2, L_3 \subseteq \Sigma^*$ (Σ beliebig).

Beweis:

a) Direkter Beweis. Angenommen $L^2 = L$. Dann gilt: $L^3 \overset{\text{Def}}{=} L^2 \cdot L \overset{\text{Annahme}}{=} L \cdot L \overset{\text{Def}}{=} L^2 \overset{\text{Annahme}}{=} L$ $\square$

b) Wir geben Beispiele: $L = \emptyset$, $L = \{\lambda\}$, $L = \Sigma^*$

c) Wir nutzen Lemma 2.2 und konstruieren ein Gegenbeispiel:

$L_1 = \{\lambda, 0\}$, $L_2 = \{0\}$, $L_3 = \{00\}$.

Dann ist: $L_1 \cdot (L_2 \cap L_3) = \emptyset$

$L_1 \cdot L_2 \cap L_1L_3 = \{0, 00\} \cap \{00, 000\} = \{00\}$

d) Nutze Lemma 2.2: $L_2 \cdot (L_1 - L_3) = L_2(L_1 \cap L_3^c) \subseteq L_2L_1 \cap L_2L_3^c = L_2L_1 - (L_2^cL_3)$

$\Rightarrow$ Aussage ist falsch.

Wir geben ein Gegenbeispiel: Sei $L_3 = \{\lambda\}$, $L_1 = \{a\}^* = L_2$. Dann ist $L - L_1 = \{a\}^+$, also $L_2 \cdot (L_1 - L_3) = \{a\}^* \cdot \{a\}^+ = \{a\}^+$. Andererseits ist $L_1L_2 = \{a\}^*\{a\}^* = \{a\}^*$ und $L_2 \cdot L_3 = \{a\}^*\{\lambda\} = \{a\}^*$. Also $L_1L_2 - L_2L_1 = \{a\}^* - \{a\}^* = \emptyset$.

Mehr Kombinatorik :

Aufgabe: Sei $\Sigma := \{x,y\}$, $n \in \mathbb{N}$. Bestimme die Anzahl der Wörter der Länge n über Σ, welche nicht das Teilwort yx enthalten. (Sollte es weitere Einschränkungen zu den Wörtern geben, so kann man dieses Resultat als Ausgangspunkt nehmen)

Beweis:

Wenn ein Wort der Länge n nicht das Teilwort yx enthält, dann hat es die Form $x^l y^m$ für irgendwelche $l, m \in \mathbb{N}$ s.d. $l + m = n$. Für gegebenes n gilt also: $0 \leq l \leq n$ und m ist durch die Wahl von l eindeutig bestimmt. Da es $n+1$ Möglichkeiten zur Wahl von l gibt, gibt es genau $n+1$ solche Wörter der Länge n. (Menge: $\{1^l 0^m \mid l, m \in \mathbb{N}, l+m=n\}$) $\square$

Einschätzung zu Serie 1:

A1| **a)** gute Übung

b) denk nicht mal drüber nach

c) gute Übung

A2| **a)** viel zu einfach für eine Klausur

b) gute Übung

A3| Meine Meinung: vollkommen irrelevant