University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Theoretische Informatik
Formale Sprachen

Prof. Dr. Juho Mäkiö
16.10. 2018

Hier geht es um….

- Verarbeitung von Sprachen
  - Sprache = Formulierung von Probleminstanzen
  - Sprache = Träger von Information
  - Sprache = Beschreibung von Systemverhalten

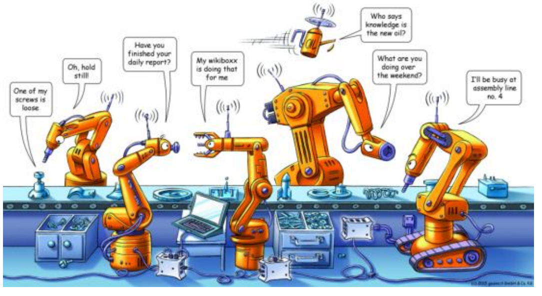
Welche Sprachen kennen Sie?

Was ist eine Sprache?

Sprache besteht aus Sätzen, die durch Verkettung aus einem Zeichenvorrat, den zur Verfügung stehenden Symbolen gebildet werden können.

Motivation – Verarbeitung von Sprachen

- Sprache bildet die Grundlage für Kommunikation
  - Mensch – Mensch
  - Mensch – Maschine
  - Maschine – Maschine



Verarbeitung von Sprachen

- Anwendung:
  - Compilerbau C, C++, JAVA, ...
  - Internet HTML
  - Datenbanken SQL, XML
  - Nutzerschnittstellen SHELL
  - Dateien einlesen XML

Beschreibung von Systemverhalten

- Formale Modelle für Systeme, formale Semantik
- Beschreibung & Analyse von Phänomenen
- Studium von Operationen auf Systemen (Komposition, Verfeinerung, …)
- Anwendung:
  - Modelprüfung
    [engl. Model Checking (Fehlersuche jenseits von Testen)]
    - Vollautomatische Verifikation einer Systembeschreibung (Modell) gegen seine Spezifikation (Formel)
  - Statische Programmanalyse, -optimierung
  - Design leistungsfähiger Operationen
  - Korrektheit per Konstruktion

Grundlegendes

Alphabet:	X	endliche Menge X = { $x_1,\dots,x_n$ }
Wort:	w	endliche Folge von Buchstaben
		w : {0, …, n-1} → X
Wortlänge	|w|:	Anzahl der Buchstaben (n)
Wortmenge	X*	die Menge aller Wörter über Alphabet X
Sprache	L	eine Menge von Wörtern über X
		$L \subseteq X^*$
Satz von L		ein Wort $w \in L$

Alphabet

- Alphabet   ∑: endliche Menge von Zeichen (meist ≠ ∅)
- <u>Bsp.</u> $\sum_1 = \{0, \ldots, 9\}$, $\sum_2 = \{0, 1\}$, $\sum_3 = \{a, \ldots, z\}$

Das phönizische Alphabet

Wort

- Wort ω (über dem Alphabet ∑) ist eine endliche Folge von Zeichen aus ∑. (-> String/Zeichenkette)
- ω besteht aus n Zeichen:
  - n = 0   :   ε ≙ <u>leeres Wort</u>
  - n > 0   : $\omega = \omega_1 \ldots \omega_n$ mit $\omega_i \in \sum$ für i = 1, …, n

  - <u>Bsp.</u>   ∑ = {0, …, 9}
  - ω = 123

- <u>Länge eines Wortes</u> → Anzahl der Zeichen (Kardinalität)
  - |ω| = n
  - Z.B. |ε| = 0 ; |123| = 3

Konkatenation (Verkettung)

Gegeben seien: $\omega = \omega_1 \ldots \omega_n$ und $v = v_1 \ldots v_m$
$\omega v = \omega_1 \ldots \omega_n v_1 \ldots v_m$ mit $|\omega v| = n + m$ (oft auch: $\omega + v$)
(oder Operation ○: $\omega \circ v = \omega_1 \ldots \omega_n \circ v_1 \ldots v_m$)

Neutrales Element:
Es gilt: $\varepsilon\, \omega = \omega$, $\omega\, \varepsilon = \omega$
($\varepsilon \triangleq$ „neutrales Element“ - WICHTIG)

Beispiel: ab + cd = abcd (Mehrdeutigkeit von ‚+‘)
(manchmal auch * oder ○)

Wort über ∑

- $\sum^n$
- $\sum^n = \{\omega \mid \omega$ ist Wort über $\sum$ und $|\omega| = n\}$
  - Wörter der Länge n
  - $\sum^0 = \{\varepsilon\}$
  - Bsp.:
    $\sum = \{0, 1\}$
    $\sum^2 = \{00, 01, 10, 11\}$

Gegeben sei $\sum = \{a, b, c\}$ mit n=3. Bestimmen Sie die $|\sum|$.

Kleensche Hülle

- Die Kleensche Hülle eines Alphabets $\sum$ ist die Menge alle Wörter, die durch beliebige Konkatenation von Symbolen des Alphabets $\sum$ gebildet werden können, inklusive das leere Wort ε.
- Menge aller Wörter $\sum^*$
- $\sum^* = \sum^0 \cup \sum^1 \cup \sum^2 \cup \sum^3 \cup \ldots$

Positive Hülle

- Die Positive Hülle
  ε-freie Menge aller Wörter $\sum^+$
  $\sum^+ = \sum^1 \cup \sum^2 \cup \ldots$
  d.h. $\sum^+ = \sum^* \setminus \{\varepsilon\}$, $\sum^* = \sum^+ \cup \{\varepsilon\}$

Bsp. $\sum = \{0, 1\}$

$\sum^+ = \{0, 1, 00, 01, 10, 11, 000, \ldots\}$
$\sum^* = \{\varepsilon, 0, 1, 00, 01, \ldots\}$

(1 ist eine Ziffer, Zahl)
1 ist Zeichen aus $\sum$
1 ist Wort der Länge 1

Manchmal: Wort \`1\` (-> String)

Aufbau eines Wortes

ω^n

$n = 0$: $\omega^0 = \varepsilon$

$n = 3$: $\omega^3 = \omega\omega\omega$

rekursiv: $\omega^0 = \varepsilon$

$\omega^n = \omega\omega^{n-1}$

Oft werden auch (,) verwendet:

$1\ (01)^3\ 1 = 1\ 010101\ 1$

Die Klammern (,) dürfen nicht zu $\sum$ gehören

Spiegelung eines Wortes ω^R

- $\omega = \omega_1 \dots \omega_n$: $\omega^R = \omega_n \dots \omega_1$ mit $\omega_i \in \sum$

  Bsp.

  $\varepsilon^R = \varepsilon$

  $a^R = a$

  $\omega^R = ba$ mit $\omega = ab$

  $(abc)^R = cba$

  $\omega = \omega^R$ → Palindrom

  Bsp. OTTO

  RELIEFPFEILER

  LAGERREGAL

Formale Sprache

Eine (formale) Sprache L über einem Alphabet Σ ist eine Menge von Wörtern über Σ: $L \subset \Sigma^*$

<u>Bsp.</u>

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

$L_1 = \{0, 1, 00, 01, 10, 11\}$

$L_2 = \{1, 10, 100, 1000, \ldots\}$

L bzw. die Elemente von L (Wörter) haben zunächst keine Semantik/Bedeutung.

Beschreibung von formalen Sprachen

Angabe einer Eigenschaft

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

$L_1 = \{\omega \in \Sigma^* \mid \omega \text{ beginnt mit } 1\}$

$L_1 = \{1, 10, 11, 100, \ldots\}$

$L_2 = \{\omega \in \Sigma^* \mid |\omega| = 3\}$

$L_2 = \{000, 001, \ldots, 111\}$

Angabe eines Konstruktionsmusters

$L_3 = \{1\omega \mid \omega \in \Sigma^*\}$

$L_3 = \{1, 10, 11, \dots\} = L_1$

$L_4 = \{v1\omega \mid v, \omega \in \Sigma^*\}$

$L_4 = \{1, 01, 010, \dots\}$

$L_4 = \{\omega \in \Sigma^* \mid \omega \text{ enthält mindestens eine } 1\}$

Bedingte Kardinalität

$b \in \Sigma, \quad \omega \in \Sigma^* \quad : \quad |\omega|_b \quad \triangleq$ Anzahl der in ω vorkommenden b's

Bsp.

$|abac|_a = 2$

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

$L_1 = \{(0^n 1^n)^n \mid n \in \mathbb{N}\}$

$L_1 = \{\varepsilon, 01, 00110011, \dots\}$

$L_2 = \{1\omega 0 \mid \omega \in \Sigma^* \wedge |\omega|_1 = 2\}$

$L_2 = \{1110, 101100, \dots\}$

Beispiel

<u>Bsp.</u> $\sum = \{0, 1\}$

a)

$L_1 = \{\omega 1 v \mid \omega, v \in \sum^*\}$

Konstruktionsmuster

$L_1 = \{\omega \mid \omega \in \sum^* \wedge |\omega|_1 \geq 1\}$

$L_1 = \{\omega \mid \omega \in \sum^* \wedge \omega$ enthält mindestens eine 1}

Verschiedene Beschreibungen für dieselbe Sprache.

Beispiel

$L_2 = \{x1x \mid x \in \sum\} = \{010, 111\}$

$L_3 = \{\omega \mid \omega \in \sum^* \wedge |\omega|_1 = 1\}$

$L_3 = \{1, 01, 010, 001, \ldots\}$

$L_3 = \{\underbrace{0^n 1 0^m}_{\text{Konstruktionsmuster}} \mid n, m \in \mathbb{N}; n, m \geq 0\}$

$L_4 = \{\omega \mid \omega \in \sum^* \wedge \omega$ enthält genau 2 Einser}

$L_4 = \{11, 011, 0101, 010010, \ldots\}$

Yes we can…

a_1
a_2
a_3

- $\mathbb{N}$ $\mathbb{Z}$ $\mathbb{R}$ $\mathbb{Q}$