

**Vorlesung**

# **Grundlagen der Theoretischen Informatik / Einführung in die Theoretische Informatik I**

**Bernhard Beckert**

**Institut für Informatik**



**Sommersemester 2007**

Diese Vorlesungsmaterialien basieren ganz wesentlich auf den Folien zu den Vorlesungen von

**Katrin Erk** (gehalten an der Universität Koblenz-Landau)

**Jürgen Dix** (gehalten an der TU Clausthal)

Ihnen beiden gilt mein herzlicher Dank.

*– Bernhard Beckert, April 2007*

## Inhalt von Teil II

In den folgenden Abschnitten führen wir die Begriffe

- **Sprache**
- **Grammatik**

ein.

Wir untersuchen insbesondere

- 1 wie man Probleme aus der Mathematik, Graphentheorie, Logik als **Probleme über Sprachen** formulieren kann.
- 2 wie man Klassen von Grammatiken von steigendem Schwierigkeitsgrad definiert: **Chomsky-Hierarchie**.
- 3 **wieviele** Grammatiken und Sprachen **es überhaupt gibt** (soviele wie natürliche Zahlen, reelle Zahlen oder komplexe Zahlen?)

## Teil II

# Terminologie

- 1 **Sprache, Grammatik**
- 2 Warum Sprachen?
- 3 Die Chomsky-Hierarchie
- 4 Probleme über Sprachen
- 5 Endlich, unendlich und dann?

## Definition 6.1 (Alphabet)

Ein **Alphabet** ist eine Menge von Zeichen/Buchstaben

- Grundlage einer Sprache (die zur Verfügung stehenden Zeichen)
- Meist endlich

## Definition 6.2 (Wort)

Ein **Wort** (über einem Alphabet  $\Sigma$ )  
ist eine endliche Folge von Zeichen aus  $\Sigma$

$|w|$  bezeichnet Länge eines Wortes  $w$

$\varepsilon$  bezeichnet das **leere Wort**

## Operationen auf Wörtern

### Verknüpfung (Konkatenation):

$$w \circ w'$$

assoziativ, oft geschrieben als  $ww'$

### $i$ -te Potenz:

$$w^0 = \varepsilon, \quad w^{i+1} = ww^i$$

### Reverse:

$w^R$  = das Wort  $w$  rückwärts

## Definition 6.3 (Sprache)

Eine Sprache  $L$  (über einem Alphabet  $\Sigma$ )  
ist eine Menge von Wörtern über  $\Sigma$ .

## Operationen auf Sprachen

### Konkatenation:

$$L \circ M = \{w \circ w' \mid w \in L, w' \in M\}$$

### $i$ -te Potenz:

$$L^0 = \{\varepsilon\}, \quad L^{i+1} := LL^i$$

### Reverse:

$$L^R = \{w^R : w \in L\}$$

## Kleene-Hülle

$$L^* = L^0 \cup L^1 \cup L^2 \cup \dots$$

Variante:

$$L^+ = LL^* = L^1 \cup L^2 \cup \dots$$

$\Sigma^*$  bezeichnet die Menge aller Wörter über  $\Sigma$

## Genau genommen besteht ein Unterschied:

ein Buchstabe  $\neq$  Wort, das nur aus dem einen Buchstaben besteht

Darum ist  $\Sigma$  selbst keine Sprache über  $\Sigma$

(Oft wird über diesen Unterschied hinweggesehen)

## Definition 6.4 (Reguläre Ausdrücke)

Menge  $\mathfrak{Reg}_\Sigma$  der **regulären Ausdrücke** (über  $\Sigma$ ) ist definiert durch:

- 1 0 ist ein regulärer Ausdruck
- 2 Für jedes  $a \in \Sigma$  ist  $a$  ein regulärer Ausdruck
- 3 Sind  $r$  und  $s$  reguläre Ausdrücke, so auch
  - $(r + s)$  (Vereinigung),
  - $(rs)$  (Konkatenation),
  - $(r^*)$  (Kleene Stern)

Klammern können weggelassen werden, dann

- \* hat Vorrang vor Konkatenation
- Konkatenation hat Vorrang vor +

# Reguläre Ausdrücke

## Definition 6.5 (Semantik regulärer Ausdrücke)

Ein regulärer Ausdruck  $r$  stellt eine Sprache  $\mathcal{J}(r)$  über  $\Sigma$  wie folgt dar:

$$\mathcal{J}(0) := \emptyset$$

$$\mathcal{J}(a) := \{a\} \quad \text{für } a \in \Sigma$$

$$\mathcal{J}(r+s) := \mathcal{J}(r) \cup \mathcal{J}(s)$$

$$\mathcal{J}(rs) := \mathcal{J}(r)\mathcal{J}(s)$$

$$\mathcal{J}(r^*) := \mathcal{J}(r)^*$$

## Wir benutzen auch das Makro ...

$$1 := 0^*$$

Es gilt:  $\mathcal{J}(1) = \{\varepsilon\}$

## Übung

Welche Sprachen werden durch die folgenden regulären Ausdrücke dargestellt?

- $aa$
- $(a + b)^*$
- $aa^* + bb^*$