Vorlesung

Grundlagen der Theoretischen Informatik / Einführung in die Theoretische Informatik I

Bernhard Beckert

Institut für Informatik

UNIVERSITÄT KOBLENZ · LANDAU

Sommersemester 2007

Dank

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

Alphabete, Wörter

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 Σ)
ist eine endliche Folge von Zeichen aus Σ

$|w|$ bezeichnet Länge eines Wortes w

ε bezeichnet das leere Wort

Alphabete, Wörter

Operationen auf Wörtern

Verknüpfung (Konkatenation):

$w \circ w'$

assoziativ, oft geschrieben als ww'

i-te Potenz:

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

Reverse:

$w^R =$ das Wort w rückwärts

Sprache

Definition 6.3 (Sprache)

Eine Sprache L (über einem Alphabet Σ)
ist eine Menge von Wörtern über Σ.

Operationen auf Sprachen

Konkatenation:

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

i-te Potenz:

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

Reverse:

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

Sprache

Kleene-Hülle

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

Variante:

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

Σ^* bezeichnet die Menge aller Wörter über Σ

Genau genommen besteht ein Unterschied:

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

Darum ist Σ selbst keine Sprache über Σ

(Oft wird über diesen Unterschied hinweggesehen)

Reguläre Ausdrücke

Definition 6.4 (Reguläre Ausdrücke)

Menge $\mathfrak{Reg}_\Sigma$ der **regulären Ausdrücke** (über Σ) 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 $\mathfrak{I}(r)$ über Σ wie folgt dar:

$$\begin{aligned}
\mathfrak{I}(0) &:= \emptyset \\
\mathfrak{I}(a) &:= \{a\} \qquad \text{für } a \in \Sigma \\
\mathfrak{I}(r+s) &:= \mathfrak{I}(r) \cup \mathfrak{I}(s) \\
\mathfrak{I}(rs) &:= \mathfrak{I}(r)\mathfrak{I}(s) \\
\mathfrak{I}(r^*) &:= \mathfrak{I}(r)^*
\end{aligned}$$

Wir benutzen auch das Makro . . .

$$1 := 0^*$$

Es gilt: $\mathfrak{I}(1) = \{\varepsilon\}$

Reguläre Ausdrücke

Übung

Welche Sprachen werden durch die folgenden regulären Ausdrücke dargestellt?

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