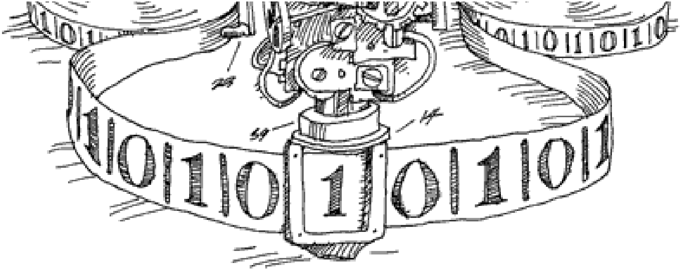
Theoretische Grundlagen der Informatik

Tutorium 10

Institut für Theoretische Informatik



Grammatiken

Definition

Eine Grammatik ist ein Regelsystem, mit dem sich die Wörter einer Sprache erzeugen lassen. Sie besteht aus vier Komponenten:

- ein endliches **Alphabet** Σ (auch Terminale genannt)
- eine endliche Menge V mit $V \cap \Sigma = \emptyset$ von **Variablen** (auch Nichtterminale genannt)
- ein **Startsymbol** $S \in V$
- eine endliche Menge von **Ableitungsregeln** R (auch Produktionen genannt). Eine Ableitungsregel ist ein Paar (l, r), wobei $l \in (V \cup \Sigma)^+$ und $r \in (V \cup \Sigma)^*$ ist. Wir schreiben meist $l \to r$.

Beispiel

Beispiel: Sprache der Palindrome

$$
\begin{aligned}
V &= \{S\} \\
\Sigma &= \{0, 1\} \\
R &= \{S \to \varepsilon \mid 0 \mid 1, \\
&\quad\; S \to 0S0 \mid 1S1\}
\end{aligned}
$$

Chomsky-Hierarchie

Von Noam Chomsky definiert, um natürliche Sprachen formalisieren zu können.
Wir definieren Klassen von Sprachen, die von Grammatiken mit bestimmten Einschränkungen erzeugt werden:

- CH-3 Regulär/Rechtslinear
- CH-2 Kontextfrei
- CH-1 Kontextsensitiv/Längenbeschränkt
- CH-0 Beliebig

Es gilt:

$$CH\text{-}3 \subset CH\text{-}2 \subset CH\text{-}1 \subset CH\text{-}0$$

Frage: Welche Sprachen liegen nicht in CH-0 ?

Chomsky-Hierarchie

Von Noam Chomsky definiert, um natürliche Sprachen formalisieren zu können.
Wir definieren Klassen von Sprachen, die von Grammatiken mit bestimmten Einschränkungen erzeugt werden:

- CH-3 Regulär/Rechtslinear
- CH-2 Kontextfrei
- CH-1 Kontextsensitiv/Längenbeschränkt
- CH-0 Beliebig (rekursiv aufzählbar)

Es gilt:

$$CH\text{-}3 \subset CH\text{-}2 \subset CH\text{-}1 \subset CH\text{-}0$$

Frage: Welche Sprachen liegen nicht in CH-0 ?

CH-0

Rekursiv aufzählbare Sprachen

- Sprachen, die von einer beliebigen Grammatik erzeugt werden
- Genau die Sprachen, die DTM & NTM *akzeptieren* können

CH-1

Kontextsensitive/längenbeschränkte Sprachen

- Sprachen, die von Grammatiken erzeugt werden, deren Ableitungen eine linke Seite besitzen, die höchstens so lang ist wie ihre rechte Seite. D.h. Produktionen machen das Wort nie kürzer.
  $\Rightarrow$ entscheidbar
- (Genau die Sprachen, die von linear beschränken Turingmaschinen erkannt werden)
- $\rightarrow$ vgl. Klasse $\mathcal{NTAPE}$(n)

Beispiel

- $\{a^n b^n c^n \mid n \in \mathbb{N}\}$
- Syntaktisch korrektes C (typedef)
- Die Sprache der wohlgeformten Latex-Dokumente

CH-1

Kontextsensitive/längenbeschränkte Sprachen

- Sprachen, die von Grammatiken erzeugt werden, deren Ableitungen eine linke Seite besitzen, die höchstens so lang ist wie ihre rechte Seite. D.h. Produktionen machen das Wort nie kürzer.
  $\Rightarrow$ entscheidbar
- (Genau die Sprachen, die von linear beschränken Turingmaschinen erkannt werden)
- $\rightarrow$ vgl. Klasse $\mathcal{NTAPE}$(n)

Beispiel

- $\{a^n b^n c^n \mid n \in \mathbb{N}\}$
- Syntaktisch korrektes C (typedef)
- Die Sprache der wohlgeformten Latex-Dokumente

CH-1

Kontextsensitive/längenbeschränkte Sprachen

- Sprachen, die von Grammatiken erzeugt werden, deren Ableitungen eine linke Seite besitzen, die höchstens so lang ist wie ihre rechte Seite. D.h. Produktionen machen das Wort nie kürzer.
  $\Rightarrow$ entscheidbar
- (Genau die Sprachen, die von linear beschränken Turingmaschinen erkannt werden)
- $\rightarrow$ vgl. Klasse $\mathcal{NTAPE}$(n)

Beispiel

- $\{a^n b^n c^n \mid n \in \mathbb{N}\}$
- Syntaktisch korrektes C (typedef)
- Die Sprache der wohlgeformten Latex-Dokumente

CH-2

Kontextfreie Sprachen

- Grammatiken, deren Ableitungsregeln auf der linken Seite immer aus genau einem Nichtterminalsymbol bestehen
- Genau die Sprachen, die von Kellerautomaten erkannt werden (Vorgriff)

Beispiele

- $\{a^n b^n \mid n \in \mathbb{N}\}$
- Viele natürlichen Sprachen
- Die Sprache der gültigen Klammerausdrücke
- Die Sprache der regulären Ausdrücke

CH-3

Reguläre Sprachen

- Grammatiken, deren Ableitungsregeln ausschließlich die folgende Form haben:

$$A \rightarrow v \text{ mit } A \in V \text{ und } v = \varepsilon \text{ oder } v = aB \text{ mit } a \in \Sigma, B \in V$$

- Genau die Sprachen, die von endlichen Automaten erkannt werden

Beispiele

- Alle endlichen Sprachen
- Tokens ($\rightarrow$ vgl. Compilerbau)
- Die Sprache der kontextfreien Grammatiken über festem Alphabet
- Perl Regexes *nicht*

Aufgabe

Karlsruher Institut für Technologie

Gegeben sei die Grammatik $G = (\Sigma, V, S, R)$ mit $\Sigma = \{a, b\}$, $V = \{S, A\}$ und

$$R = \{S \to AA, \quad A \to AAA, \quad A \to bA, \quad A \to Ab, \quad A \to a\}.$$

- Welchen Typ in der Chomsky-Hierarchie hat G?
- Welche Zeichenketten aus Σ^* können in vier Schritten abgeleitet werden?
- Sei $\alpha = (b^* ab^* ab^*)^+$ ein regulärer Ausdruck. Zeige: $L(\alpha) \subseteq L(G)$.

Aufgabe: CH-1-Grammatiken

Konstruiere eine Typ-1-Grammatik, welche die Sprache

$$L = \{a^n b^n c^n \mid n \geq 1\}$$

erzeugt.

Aufgabe: CH-1-Grammatiken

Konstruiere eine Typ-1-Grammatik, welche die Sprache

$$L = \{a^n b^n c^n \mid n \geq 1\}$$

erzeugt.

Lösung

$$\Sigma = \{a, b, c\}$$
$$V = \{S, S', \#, B, C\}$$
$$S \rightarrow S'$$
$$S' \rightarrow aS'\#C \mid aBC$$
$$C\# \rightarrow \#C$$
$$B\# \rightarrow BB$$
$$B \rightarrow b$$
$$C \rightarrow c$$

Aufgabe: CH-2-Grammatiken

Gib eine kontextfreie Grammatik an, die die Sprache

$$L = \{w \in \Sigma^* \mid w \textit{ enthält mehr Nullen als Einsen}\}$$

über dem Alphabet $\Sigma = \{0, 1\}$ beschreibt.

Aufgabe: CH-2-Grammatiken

Gib eine kontextfreie Grammatik an, die die Sprache

$$L = \{w \in \Sigma^* \mid w \textit{ enthält mehr Nullen als Einsen}\}$$

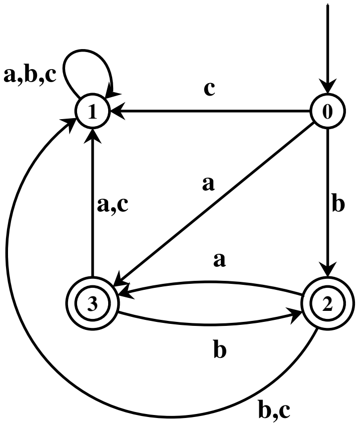
über dem Alphabet $\Sigma = \{0, 1\}$ beschreibt.

Lösung

$$\begin{aligned}
\Sigma &= \{0, 1\} \\
V &= \{S, \#\} \\
S &\to \#0\# \\
\# &\to 0\#1 \mid 1\#0 \mid \#\# \mid 0\# \mid \varepsilon
\end{aligned}$$

Aufgabe: CH-3-Grammatiken

Gib unter Anwendung des Verfahrens aus Satz 5.5 eine Grammatik an, welche genau die Sprache erzeugt, die folgender DEA akzeptiert:



Aufgabe

1. Bestimme eine möglichst kurze Grammatik G_R für reguläre Ausdrücke über dem Alphabet $\Sigma = \{0, 1\}$ mit maximalem Chomsky-Typ.
2. Bestimme einen Ableitungsbaum für das Wort $1^* \cup (01)^*$.
3. Ist die Grammatik eindeutig? Diskutiere, ob es eine eindeutige Grammatik für reguläre Ausdrücke geben kann, falls die angegebene Grammatik mehrdeutig ist.

Aufgabe

1. Bestimme eine möglichst kurze Grammatik G_R für reguläre Ausdrücke über dem Alphabet $\Sigma = \{0, 1\}$ mit maximalem Chomsky-Typ.
2. Bestimme einen Ableitungsbaum für das Wort $1^* \cup (01)^*$.
3. Ist die Grammatik eindeutig? Diskutiere, ob es eine eindeutige Grammatik für reguläre Ausdrücke geben kann, falls die angegebene Grammatik mehrdeutig ist.

Lösung

$$\Sigma = \{0, 1, {}^*, \cup, \epsilon, \emptyset, (,)\}$$
$$V = \{S, P, K\}$$

$$S \to P \cup S \mid P$$
$$P \to KP \mid K$$
$$K \to 0 \mid 1 \mid \epsilon \mid \emptyset \mid T^* \mid (S)$$

Probleme

Karlsruher Institut für Technologie

Es gibt einige interessante Probleme, mit deren Entscheidbarkeit, Komplexität bzw. Abgeschlossenheit wir uns noch befassen (jeweils für die Klassen)

- Das Wortproblem
- Vereinigung
- Schnitt
- Komplement
- Konkatenation
- Kleenescher Abschluss (*-Operator)

Lizenzen