Theoretische Informatik

Teil 4
Kontextfreie Grammatiken

Frühlimgssemester 2019

L. Di Caro
D. Flumini
O. Stern

Lernziele

- Die Studierenden können **kontextfreie Grammatiken (KFG)** formal und an Beispielen erklären. Sie können für einfache gegebene Sprachen eine zugehörige Grammatik entwerfen.
- Die Studierenden kennen **Ableitungen** für KFG, können Beispiele dazu angeben und den Begriff der **Mehrdeutigkeit** einer KFG erklären.
- Sie können die **Sprache einer KFG** definieren und den Zusammenhang zu den regulären Ausdrücken (RA) angeben.

Lerninhalte

- Einführung und Definition
- Ableitungsschritt, Ableitung
- Rechtseitige und linksseitige Ableitungen
- Sprache eine KFG
- Ableitungsbaum
- Mehrdeutigkeit (von KFG)
- Zusammenhang von KFG und regulären Ausdrücken
- Anwendungen

Einführendes Beispiel

Motivation:

- Es wurde bereits gezeigt, dass es nicht-reguläre Sprachen gibt, d. h. Sprachen, die mächtiger als RA sind bzw. sich nicht durch einen RA beschreiben lassen.
- $\{ 0^n 1^n \mid n \in \mathbb{N} \}$ ist ein Beispiel für eine einfache nicht-reguläre Sprache (vgl. Abschnitt EA).

Einführendes Beispiel

Motivation:

- Es wurde bereits gezeigt, dass es nicht-reguläre Sprachen gibt, d. h. Sprachen, die mächtiger als RA sind bzw. sich nicht durch einen RA beschreiben lassen.
- $\{ 0^n 1^n \mid n \in \mathbb{N} \}$ ist ein Beispiel für eine einfache nicht-reguläre Sprache (vgl. Abschnitt EA).

=> Dennoch lässt sich diese Sprache bzw. Vorschrift zur Erzeugung eigentlich leicht beschreiben.

Wie könnte eine solche Beschreibung informell lauten?

Einführendes Beispiel

Beispiel

Kontextfreie Grammatik für die nicht-reguläre Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

Einführendes Beispiel

Beispiel

Kontextfreie Grammatik für die nicht-reguläre Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

$$A \to 0A1$$
$$A \to \varepsilon$$

Einführendes Beispiel

Beispiel

Kontextfreie Grammatik für die nicht-reguläre Sprache $\{ 0^n 1^n \mid n \in \mathbb{N} \}$:

$$A \to 0A1$$
$$A \to \varepsilon$$

Eine Ableitung des Wortes 000111 in der Grammatik:

Einführendes Beispiel

Beispiel

Kontextfreie Grammatik für die nicht-reguläre Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

$$A \to 0A1$$
$$A \to \varepsilon$$

Eine Ableitung des Wortes 000111 in der Grammatik:

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000111$$

Einführendes Beispiel

Beispiel

Kontextfreie Grammatik für die nicht-reguläre Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

$$
\begin{aligned}
A &\rightarrow 0A1 \\
A &\rightarrow \varepsilon
\end{aligned}
$$

Eine Ableitung des Wortes 000111 in der Grammatik:

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000111$$

Anmerkung: Alle Wörter, die wir in der Grammatik ableiten können, bilden die Sprache der Grammatik.

Definition einer kontextfreien Grammatik

Definition (Kontextfreie Grammatik)

Eine **kontextfreie Grammatik** G (KFG) ist ein 4-Tupel (N, Σ, P, S) mit

- N ist das Alphabet der **Nichtterminale** (Variablen).
- Σ ist das Alphabet der **Terminale**.
- P ist eine endliche Menge von **Produktionen** (Regeln). Jede Produktion hat die Form

$$X \to \beta$$

  mit **Kopf** $X \in N$ und **Rumpf** $\beta \in (N \cup \Sigma)^*$.
- S ist das **Startsymbol**, wobei $S \in N$.

Definition (Satzform)

Ein Wort $\beta \in (N \cup \Sigma)^*$ nennen wir **Satzform**.

Beispiel einer kontextfreien Grammatik

Beispiel

Eine KFG G_1 für die Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

Beispiel einer kontextfreien Grammatik

Beispiel

Eine KFG G_1 für die Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

$$G_1 = (\{A\}, \{0, 1\}, P, A)$$

mit

$$P = \{A \to 0A1, A \to \varepsilon\}$$

Beispiel einer kontextfreien Grammatik

Beispiel

Eine KFG G_1 für die Sprache $\{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$:

$$G_1 = (\{A\}, \{0, 1\}, P, A)$$

mit

$$P = \{A \to 0A1, A \to \varepsilon\}$$

Anmerkung:Mehrere Regeln mit dem gleichen Nichtterminal im Kopf können wir kompakt notieren:

$$A \to 0A1 \mid \varepsilon$$

Grammatik für balancierte Klammerausdrücke

Beispiel

Eine KFG G_2 für die Sprache der balancierten Klammerausdrücke wie zum Beispiel $(())()$:

Grammatik für balancierte Klammerausdrücke

Beispiel

Eine KFG G_2 für die Sprache der balancierten Klammerausdrücke wie zum Beispiel $(())()$:

$$G_2 = (\{A\}, \{(,)\}, P, A)$$

Die Menge der Produktionen P ist

$$A \to (A) \mid AA \mid \varepsilon$$

Grammatik für balancierte Klammerausdrücke

Beispiel

Eine KFG G_2 für die Sprache der balancierten Klammerausdrücke wie zum Beispiel $(())()$:

$$G_2 = (\{A\}, \{(,)\}, P, A)$$

Die Menge der Produktionen P ist

$$A \to (A) \mid AA \mid \varepsilon$$

Beispiel (Ableitung des Wortes $(())()$ in G)

Grammatik für balancierte Klammerausdrücke

Beispiel

Eine KFG G_2 für die Sprache der balancierten Klammerausdrücke wie zum Beispiel $(())()$:

$$G_2 = (\{A\}, \{(,)\}, P, A)$$

Die Menge der Produktionen P ist

$$A \to (A) \mid AA \mid \varepsilon$$

Beispiel (Ableitung des Wortes $(())()$ in G)

$$A \Rightarrow AA \Rightarrow (A)A \Rightarrow (A)(A) \Rightarrow ((A))(A) \Rightarrow (())(A) \Rightarrow (())()$$

Ableitung

Definition

Seien α, β und γ Satzformen und $A \to \gamma$ eine Produktion.

- Durch einen **Ableitungsschritt** wird eine Satzform $\alpha A \beta$ durch die Anwendung der Produktion $A \to \gamma$ in die Satzform $\alpha\gamma\beta$ **abgeleitet**. Das notieren wir mit

$$\alpha A \beta \Rightarrow \alpha\gamma\beta$$

- Eine **Ableitung** ist eine Folge von Ableitungsschritten, so dass aus einer Satzform α das Wort w abgeleitet wird.

$$\alpha \Rightarrow \ldots \Rightarrow w$$

Linksseitige- und rechtsseitige Ableitung

Definition

- Eine **linksseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *links* in der Satzform auftritt.
- Eine **rechtsseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *rechts* in der Satzform auftritt.

Linksseitige- und rechtsseitige Ableitung

Definition

- Eine **linksseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *links* in der Satzform auftritt.
- Eine **rechtsseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *rechts* in der Satzform auftritt.

Beispiel (Linksseitige Ableitung)

$$A \Rightarrow AA \Rightarrow$$

Linksseitige- und rechtsseitige Ableitung

Definition

- Eine **linksseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *links* in der Satzform auftritt.
- Eine **rechtsseitige Ableitung** ersetzt bei jedem Ableitungsschritt das Nichtterminal, das am weitesten *rechts* in der Satzform auftritt.

Beispiel (Linksseitige Ableitung)

$$A \Rightarrow AA \Rightarrow (A)A \Rightarrow ((A))A \Rightarrow (())A \Rightarrow (())(A) \Rightarrow (())()$$

Ableitung

Definition

Ein Wort $w \in \Sigma^*$ ist in einer kontextfreien Grammatik $G = (N, \Sigma, P, S)$ **ableitbar**, falls es eine Ableitung in G gibt, die mit dem Startsymbol S beginnt und mit dem Wort w endet. Dafür schreiben wir

$$S \stackrel{*}{\Rightarrow} w$$

Wir sagen auch, dass w von S ***erzeugt*** oder ***generiert*** wird.

Kontextfreie Sprachen

Definition (Sprache einer kontextfreien Grammatik)

Die von G **erzeugte Sprache** $L(G)$ beinhaltet alle Wörter, die in G aus dem Startsymbol S ableitbar sind.

$$L(G) = \{\, w \in \Sigma^* \mid S \overset{*}{\Rightarrow} w \,\}$$

Kontextfreie Sprachen

Definition (Sprache einer kontextfreien Grammatik)

Die von G **erzeugte Sprache** $L(G)$ beinhaltet alle Wörter, die in G aus dem Startsymbol S ableitbar sind.

$$L(G) = \{\, w \in \Sigma^* \mid S \overset{*}{\Rightarrow} w \,\}$$

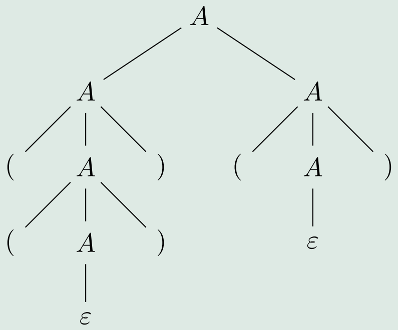
Definition (Kontextfreie Sprache)

Wenn es für eine Sprache L eine kontextfreie Grammatik G gibt mit $L = L(G)$, dann nennen wir L eine **kontextfreie Sprache**.

Ableitungsbaum

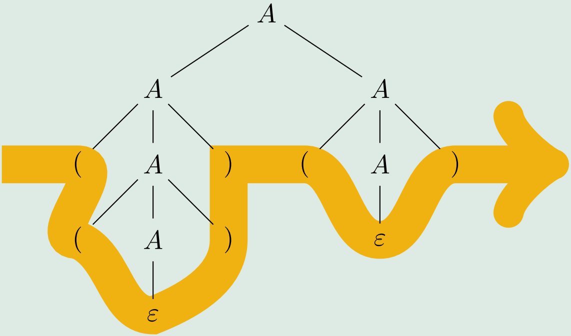
Ein **Ableitungsbaum** ist eine graphische Darstellung einer Ableitung.

Beispiel (Ableitungsbaum für das Wort $(())()$ in G_2)



Ein **Ableitungsbaum** ist eine graphische Darstellung einer Ableitung.

Beispiel (Ableitungsbaum für das Wort $(())()$ in G_2)



Ableitungsbaum

Anmerkungen:

- Ableitungsbäume werden typischerweise von Compilern als Datenstruktur für die interne Repäsentation von Quellprogrammen erzeugt.
- Sie werden auch als *Parsebäume* bezeichnet.
- Sie verdeutlichen, wie Symbole der terminalen Zeichenreihen bzw. Satzformen einer KFG in Teilzeichenreihen strukturiert werden.

Mehrdeutigkeit kontextfreier Grammatiken

Beispiel (Kontextfreie Grammatik für arithmetische Ausdrücke)

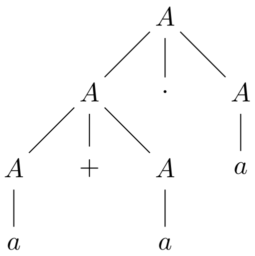
$$A \to A + A \qquad A \to A \cdot A \qquad A \to (A) \qquad A \to a$$

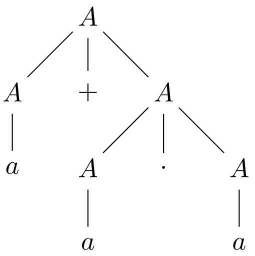
Mehrdeutigkeit kontextfreier Grammatiken

Beispiel (Kontextfreie Grammatik für arithmetische Ausdrücke)

$$A \to A + A \qquad A \to A \cdot A \qquad A \to (A) \qquad A \to a$$

Das Wort $a + a \cdot a$ kann durch zwei verschiedene Ableitungsbäume dargestellt werden.





Mehrdeutigkeit kontextfreier Grammatiken

Ein Wort, das mehrere Ableitungsbäume besitzt, hat auch mehrere linksseitige (rechtsseitige) Ableitungen.

Beispiel

Zwei linksseitige Ableitungen für $a + a \cdot a$:

- $A \Rightarrow A \cdot A \Rightarrow A + A \cdot A \Rightarrow a + A \cdot A \Rightarrow \dots$
- $A \Rightarrow A + A \Rightarrow a + A \Rightarrow a + A \cdot A \Rightarrow \dots$

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Mehrdeutige kontextfreie Grammatik)

Eine kontextfreie Grammatik nennen wir **mehrdeutig**, wenn es ein Wort gibt, das mehrere Ableitungsbäume besitzt.

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Mehrdeutige kontextfreie Grammatik)

Eine kontextfreie Grammatik nennen wir **mehrdeutig**, wenn es ein Wort gibt, das mehrere Ableitungsbäume besitzt.

Möglichkeiten, die Mehrdeutigkeit von KFG zu eliminieren:

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Mehrdeutige kontextfreie Grammatik)

Eine kontextfreie Grammatik nennen wir **mehrdeutig**, wenn es ein Wort gibt, das mehrere Ableitungsbäume besitzt.

Möglichkeiten, die Mehrdeutigkeit von KFG zu eliminieren:

- Korrekte Klammerung vom Benutzer erzwingen:
  z. B. $a + (a \cdot a)$ statt $a + a \cdot a$

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Mehrdeutige kontextfreie Grammatik)

Eine kontextfreie Grammatik nennen wir **mehrdeutig**, wenn es ein Wort gibt, das mehrere Ableitungsbäume besitzt.

Möglichkeiten, die Mehrdeutigkeit von KFG zu eliminieren:

- Korrekte Klammerung vom Benutzer erzwingen:
  z. B. $a + (a \cdot a)$ statt $a + a \cdot a$
- Grammatik anpassen (z. B. "Terme", "Faktoren")

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Mehrdeutige kontextfreie Grammatik)

Eine kontextfreie Grammatik nennen wir **mehrdeutig**, wenn es ein Wort gibt, das mehrere Ableitungsbäume besitzt.

Möglichkeiten, die Mehrdeutigkeit von KFG zu eliminieren:

- Korrekte Klammerung vom Benutzer erzwingen:
  z. B. $a + (a \cdot a)$ statt $a + a \cdot a$
- Grammatik anpassen (z. B. "Terme", "Faktoren")
- Den Produktionen einen Vorrang vergeben:
  z. B. $A \rightarrow A + A$ den Vorrang über $A \rightarrow A \cdot A$ geben

Mehrdeutigkeit kontextfreier Grammatiken

Definition (Inhärent mehrdeutig)

Eine kontextfreie Sprache, für die alle Grammatiken mehrdeutig sind, heisst **inhärent mehrdeutig**.

Beispiel (Inhärent mehrdeutige Sprache)

$$\{\, a^i b^j c^k \mid i = j \text{ oder } j = k \,\}$$

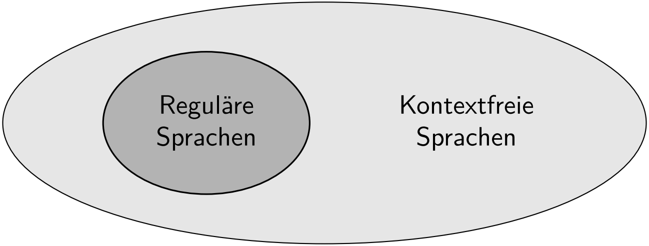
Sprach-Hierarchie

Theorem

Die kontextfreien Sprachen enthalten die regulären Sprachen.

Beweisidee.

Jede reguläre Sprache kann durch eine kontextfreie Grammatik beschrieben werden. □



KFG für eine reguläre Sprache

Sei L eine reguläre Sprache. Dann gibt es einen DEA $M = (Q, \Sigma, \delta, q_0, F)$ mit $L(M) = L$.

Dann können wir eine KFG für L wie folgt bauen:

1. Für jeden Zustand q_i gibt es ein Nichtterminal Q_i.
2. Für jede Transition $\delta(q_i, a) = q_j$ erstellen wir die Produktion $Q_i \to aQ_j$.
3. Für jeden akzeptierenden Zustand $q_i \in F$ erstellen wir die Produktion $Q_i \to \varepsilon$.
4. Das Nichtterminal Q_0 wird zum Startsymbol.

KFG für eine reguläre Sprache

Beispiel

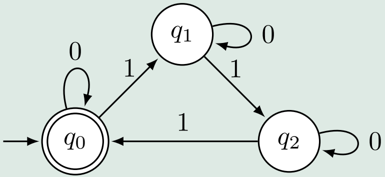
$L = \{\, w \in \{0,1\}^* \mid |w|_1 \mod 3 = 0\}$

KFG für eine reguläre Sprache

Beispiel

$L = \{\, w \in \{0,1\}^* \mid |w|_1 \mod 3 = 0\}$

Nichtterminale: Q_0, Q_1, Q_2

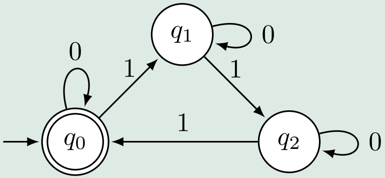


Produktionen:

KFG für eine reguläre Sprache

Beispiel

$L = \{ w \in \{0,1\}^* \mid |w|_1 \mod 3 = 0\}$
Nichtterminale: Q_0, Q_1, Q_2



Produktionen:

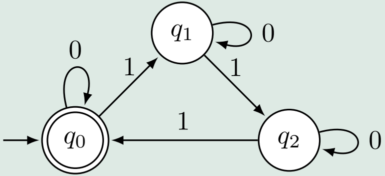
$$Q_0 \rightarrow 0Q_0 \mid 1Q_1 \mid \varepsilon$$
$$Q_1 \rightarrow 0Q_1 \mid 1Q_2$$
$$Q_2 \rightarrow 0Q_2 \mid 1Q_0$$

KFG für eine reguläre Sprache

Beispiel

$L = \{\, w \in \{0,1\}^* \mid |w|_1 \mod 3 = 0\}$

Nichtterminale: Q_0, Q_1, Q_2



Produktionen:

$$Q_0 \to 0Q_0 \mid 1Q_1 \mid \varepsilon$$
$$Q_1 \to 0Q_1 \mid 1Q_2$$
$$Q_2 \to 0Q_2 \mid 1Q_0$$

Beispiel für Ableitung von $w = 10011$:

$$Q_0 \Rightarrow 1Q_1 \Rightarrow 10Q_1 \Rightarrow 100Q_1 \Rightarrow 1001Q_2 \Rightarrow 10011Q_0 \Rightarrow 10011$$

Entwurf von kontextfreien Grammatiken

Techniken für den Entwurf von kontextfreien Grammatiken:

- Komplexe KFGs können oft in mehrere einfachere KFGs aufgeteilt werden und danach mit der Regel $S \to S_1 \mid S_2 \mid \ldots \mid S_k$ kombiniert werden.
- Um eine KFG für eine reguläre Sprache zu erstellen, kann zuerst ein DEA erstellt werden und dieser dann in eine KFG umgewandelt werden.
- Kontextfreie Sprachen enthalten manchmal Teilwörter, die voneinander „abhängig" sind. Eine KFG für diese Situation kann mit einer Regel $R \to uRv$ behandelt werden.
- Komplexere Sprachen sind meist rekursiv aufgebaut. Steht zum Beispiel das Nichtterminal A für einen Ausdruck, kann A wiederum überall dort verwendet werden, wo dieser Ausdruck erlaubt ist.

Entwurf von kontextfreien Grammatiken

Techniken für den Entwurf von kontextfreien Grammatiken:

Anmerkung: Ein häufiger Fehler beim Entwurf einer KFG besteht darin, dass neben den Wörtern einer Sprache auch Wörter erzeugt werden können, die nicht zu der Sprache gehören.

Anwendungen

- **Compiler / Parsergeneratoren**

  Teil der Syntaxanalyse: z. B. Erkennen von Programmstrukturen wie Klammern, Bedingungen (*if – else*), Blöcke (*begin – end*) ...

Anwendungen

- **Compiler / Parsergeneratoren**

  Teil der Syntaxanalyse: z. B. Erkennen von Programmstrukturen wie Klammern, Bedingungen (*if – else*), Blöcke (*begin – end*) . . .

  Frage: Wie lautet eine KFG für die Erkennung von *if – else* Blöcken, wenn *else* optional ist?

- **Compiler / Parsergeneratoren**

  Teil der Syntaxanalyse: z. B. Erkennen von Programmstrukturen wie Klammern, Bedingungen (*if – else*), Blöcke (*begin – end*) ...

  Frage: Wie lautet eine KFG für die Erkennung von *if – else* Blöcken, wenn *else* optional ist?

  Mögliche Lösung:

  $G = \{\{A\}, \{if, else\}, P, A\}$ mit $P = A \to AA \mid ifAelse \mid ifA \mid \varepsilon$

Anwendungen

- **Compiler / Parsergeneratoren**

  Teil der Syntaxanalyse: z. B. Erkennen von Programmstrukturen wie Klammern, Bedingungen (*if – else*), Blöcke (*begin – end*) . . .

  Frage: Wie lautet eine KFG für die Erkennung von *if – else* Blöcken, wenn *else* optional ist?

  Mögliche Lösung:

  $G = \{\{A\}, \{if, else\}, P, A\}$ mit $P = A \to AA \mid ifAelse \mid ifA \mid \varepsilon$

- **Markup-Sprachen** (z. B. HTML), **XML** . . .

Anmerkung: Für die Anwendung sind insbesondere nicht mehrdeutige KFGs von Interesse, da z. B. Compiler und Parser deterministisch arbeiten müssen. Ableitungsbäume stellen die interne Datenstruktur für die Repräsentation von Quellprogrammen dar.