Formale Sprachen und Komplexität
Theoretische Informatik für Medieninformatiker
Sommersemester 2022

Kontextfreie Sprachen: Greibach-Normalform und Abschlusseigenschaften

Prof. Dr. David Sabel

LFE Theoretische Informatik

Letzte Änderung der Folien: 31. Mai 2022

Greibach-Normalform

Definition (Greibach-Normalform)

Ein CFG $G = (V, \Sigma, P, S)$ mit $\varepsilon \notin L(G)$ ist in *Greibach-Normalform*, falls alle Produktionen in P von der Form $A \to aB_1B_2 \dots B_j$ mit $j \geq 0$, $A, B_1, \dots, B_j \in V$ und $a \in \Sigma$ sind.

Bemerkungen:

- benannt nach Sheila A. Greibach
- Reguläre Grammatiken sind Spezialfall der Greibach-NF: Dort ist nur $j = 0$ oder $j = 1$ erlaubt.
- Die Greibach-Normalform wird u.a. verwendet, um zu zeigen, dass Kontextfreie Sprachen genau von den nichtdeterministischen Kellerautomaten erkannt werden (später).

Algorithmus 7: Herstellen der Greibach-Normalform

Eingabe: CFG $G = (\{A_1, \dots, A_n\}, \Sigma, P, A_1)$ in Chomsky-Normalform mit $\varepsilon \notin L(G)$
Ausgabe: CFG G' in Greibach-Normalform mit $L(G) = L(G')$
Beginn
  für $i = 1$ *bis* n **tue**
    für $j = 1$ *bis* $i - 1$ **tue**
      für alle $A_i \to A_j u \in P$ **tue**
        Seien $A_j \to w_1 \mid \cdots \mid w_m$ alle Regeln in P mit A_j als linker Seite;
        Ersetze $A_i \to A_j u$ durch $A_i \to w_1 u \mid \dots \mid w_m u$ in P;
    wenn $A_i \to A_i u \in P$ **dann**
      Eliminiere die Regel mit der Operation „Elimination der Links-Rekursion";
      Sei B_i die dabei neu erzeugte Variable;
  für $i = n - 1$ *bis* 1 **tue**
    für alle $A_i \to A_j u \in P$, $j > i$ **tue**
      Seien $A_j \to w_1 \mid \cdots \mid w_m$ alle Regeln mit A_j als linker Seite;
      Ersetze $A_i \to A_j u$ durch $A_i \to w_1 u \mid \cdots \mid w_m u$ in P;
  für $i = 1$ *bis* n **tue**
    für alle $B_i \to A_j u \in P$ **tue**
      Seien $A_j \to w_1 \mid \cdots \mid w_m$ alle Regeln mit A_j als linker Seite;
      Ersetze $B_i \to A_j u$ durch $B_i \to w_1 u \mid \cdots \mid w_m u$ in P;

Ziel der geschachtelten Für-Schleife: Es gibt $A_i \to A_j u$ nur für $j > i$

Ersetzen der Regeln $A_i \to A_j u$ mit $i > j$

Ersetzen der Regeln $A_i \to A_i u$

Nun gilt für $A_i \to A_j u$ stets $j > i$.
Damit gilt für $A_n \to u$: u beginnt mit Zeichen aus Σ.
Nächste Schleife: Ersetze alle $A_i \to A_j u$ mit $j > i$

$w_1, \dots, w_m$ fangen mit Zeichen aus Σ an, da Schleife absteigend läuft!

Behandle die neuen Regeln mit B_i als linker Seite.

Korrektheit

Satz

Zu jeder CFG G mit $\varepsilon \notin L(G)$ gibt es eine CFG G' in Greibach-Normalform, sodass $L(G) = L(G')$ gilt.

Korrektheit folgt:

- durch Prüfen der genannten Invarianten
- Korrektheit der Elimination von Links-Rekursion
- Korrektheit der Operation „Inlining von Produktionen"

Ein Beispiel zur Herstellung der Greibach-Normalform ist im Skript

Abschlusseigenschaften der kontextfreien Sprachen

Theorem

Die kontextfreien Sprachen sind abgeschlossen unter Vereinigung, Produkt und Kleeneschem Abschluss.

Beweis:

- Seien L_1,L_2 CFLs und $G_i = (V_i, \Sigma_i, P_i, S_i)$ CFGs mit $L(G_i) = L_i$ für $i = 1, 2$. O.B.d.A. sei $V_1 \cap V_2 = \emptyset$.
- Seien S, S' neue Variablen ($\{S, S'\} \cap (V_1 \cup V_2) = \emptyset$).

Vereinigung: Sei

$$G_\cup = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, P_1 \cup P_2 \cup \{S \to S_1 \mid S_2\}, S)$$

Dann gilt: $L(G_\cup) = L(G_1) \cup L(G_2) = L_1 \cup L_2$.

Abschlusseigenschaften der kontextfreien Sprachen (2)

Beweis (Fortsetzung):

Produkt: Sei

$$G_\circ = (V_1 \cup V_2 \cup \{S\}, \Sigma_1 \cup \Sigma_2, P_1 \cup P_2 \cup \{S \to S_1 S_2\}, S)$$

Dann gilt $L(G_\circ) = L(G_1)L(G_2) = L_1 L_2$.

Abschlusseigenschaften der kontextfreien Sprachen (3)

Beweis (Fortsetzung):

Kleenescher Abschluss: Sei

$$G_{1,*} = (V_1 \cup \{S', S\}, \Sigma, P', S')$$

mit

$$P' = (P \setminus \{S_1 \to \varepsilon\}) \cup \{S' \to \varepsilon, S' \to S, S \to SS, S \to S_1\}$$

Dann gilt $L(G_{1,*}) = L(G_1)^*$.

Nichtkontextfreie Sprache

Satz

Die Sprache $L = \{a^l b^l c^l \mid l \in \mathbb{N}\}$ ist nicht kontextfrei.

Ein Beweis hierfür sehen wir später mit dem Pumping-Lemma für Kontextfreie Sprachen.

Wir verwenden die Sprache und die Eigenschaft jedoch im Folgenden.

Abschlusseigenschaften der kontextfreien Sprachen (4)

Theorem

Die kontextfreien Sprachen sind **nicht abgeschlossen** unter Schnitt- und Komplementbildung.

Beweis:

Schnittbildung:

- Sei $L_1 = \{a^n b^m c^m \mid m, n \in \mathbb{N}\}$ und sei $L_2 = \{a^m b^m c^n \mid m, n \in \mathbb{N}\}$
- $G_1 = (\{A, D, S\}, \{a, b, c\}, \{S \to AD, A \to \varepsilon \mid aA, D \to bDc \mid \varepsilon\}, S)$
  $G_2 = (\{C, D, S\}, \{a, b, c\}, \{S \to DC, C \to cC \mid \varepsilon, D \to aDb \mid \varepsilon\}, S)$
  Für i=1,2: $L(G_i) = L_i$, daher: L_1 und L_2 sind beide kontextfrei.
- $L_1 \cap L_2 = \{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ ist nicht kontextfrei.
- Daher sind die CFLs nicht abgeschlossen bezüglich Schnittbildung.

Abschlusseigenschaften der kontextfreien Sprachen (5)

Beweis (Fortsetzung):

Komplement:

- Widerspruchsbeweis: Wir nehmen an, es gilt:
  $$L \text{ ist CFL} \implies \overline{L} \text{ ist CFL}$$
- Seien L_1, L_2 CFLs. Dann ist auch $\overline{\overline{L_1} \cup \overline{L_2}}$ CFL
  (da CFLs abgeschlossen bez. $\bar{\cdot}$ und $\cup$)
- Aber: $\overline{\overline{L_1} \cup \overline{L_2}} = L_1 \cap L_2$. Widerspruch!