Theoretische Informatik 1

WiSe 23/24

Prof. Dr. Sebastian Siebertz
AG Theoretische Informatik
MZH, Raum 3160
siebertz@uni-bremen.de

Universität Bremen

Wiederholung kontextfreie Grammatiken

- G ist Grammatik vom *Typ 2 (kontextfrei)*, falls alle Regeln die Form $A \longrightarrow w$ haben mit $A \in N, w \in (\Sigma \cup N)^*$.
- Eine kontextfreie Grammatik (N, Σ, P, S) ist in *Chomsky-Normalform*, wenn alle Ableitungsregeln die folgende Form haben:
  - $A \longrightarrow a$ oder $A \longrightarrow BC$ mit $A, B, C \in N,\ a \in \Sigma$
  - $S \longrightarrow \varepsilon$ ist erlaubt, wenn es keine Regeln $A \to BC$ mit $S \in \{B, C\}$ gibt.
- Jede kontextfreie Grammatik lässt sich in Chomsky-Normalform umwandeln.
- Die Chomsky-Normalform ist Basis für effiziente Lösung des Wortproblems für kontextfreie Grammatiken (CYK Algorithmus).

Zusammenfassung Entscheidungsprobleme

	Wortproblem	Leerheitsprob.	Äquivalenzprob.
Typ 0			Unentscheidbar
Typ 1			Unentscheidbar
Typ 2	Polyzeit	Polyzeit	Unentscheidbar
Typ 3 NEA/reg. Ausdr.	Linearzeit	Linearzeit	Exponentialzeit
Typ 3 DEA	Linearzeit	Linearzeit	Polynomialzeit

Pumping-Lemma für kontextfreie Sprachen

Pumping-Lemma für regulära Sprachen

Wenn L eine reguläre Sprache ist, dann gilt:
es existiert $n_0 \geq 1$, so dass
für alle $w \in L$ mit $|w| \geq n_0$ gilt:
es existiert eine Zerlegung $w = xyz$ mit $y \neq \varepsilon$ und $|xy| \leq n_0$, so dass
für alle $k \geq 0$ gilt $xy^k z \in L$.

Pumping-Lemma für kontextfreie Sprachen

Wenn L eine kontextfreie Sprache ist, dann gilt:
es existiert $n_0 \geq 1$, so dass
für alle $z \in L$ mit $|z| \geq n_0$ gilt:
es existiert eine Zerl. $z = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq n_0$, so dass
für alle $k \geq 0$ gilt $uv^k wx^k y \in L$.

Beispiel

Beispiel

$L = \{a^n b^n c^n \mid n \geq 1\}$ ist nicht kontextfrei.

Beweis.

- Wir benutzen die Kontraposition des Pumping Lemmas
- Sei $n_0 \geq 1$ die Zahl aus dem Pumping Lemma.
- Betrachte $z = a^{n_0} b^{n_0} c^{n_0} \in L$.
- Wir betrachten eine beliebige Zerlegung

$$z = uvwxy, \;\; vx \neq \varepsilon \text{ und } |vwx| \leq n_0$$

Beispiel fortgesetzt

- Wir betrachten eine beliebige Zerlegung

$$z = uvwxy, \ vx \neq \varepsilon \text{ und } |vwx| \leq n_0.$$

1. Fall: v enthält verschiedene Symbole. Dann gilt

$$uv^2wx^2y \notin a^*b^*c^* \supseteq L.$$

2. Fall: x enthält verschiedene Symbole: analog.

3. Fall: v und x enthalten beide nur gleiche Symbole.

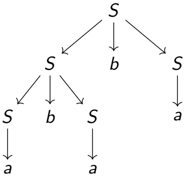
    - Dann gibt es ein Symbol $s \in \{a, b, c\}$, das in xv nicht vorkommt.
    - Dann kommt s in $uv^0wx^0y = uwy$ weiterhin n_0-mal vor.
    - Aber es gilt $|uwy| < 3n_0 \Rightarrow uwy \notin L$.

- Also ist L nicht kontextfrei.

Beweis Pumping Lemma

- Sei G eine kontextfreie Grammatik mit $L(G) = L$.
- O. B. d. A. können wir annehmen, dass $\varepsilon \notin L$ (da $n_0 \geq 1$ muss das leere Wort nicht zerlegt werden).
- Wir nehmen an, dass G in Chomsky Normalform vorliegt.
  - Keine Kettenregeln.
  - Keine Regel $A \longrightarrow \varepsilon$.
- Sei
  - m die Anzahl der Nichtterminale in G und
  - $n_0 = 2^{m+1}$.

Ableitungsbäume

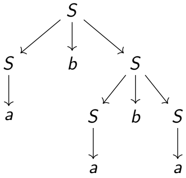


- Sei $G = (N, \Sigma, P, S)$ eine kontextfreie Grammatik.
- Ein *partieller Ableitungsbaum* ist ein gewurzelter, geordneter Baum, dessen Knoten mit Symbolen aus $N \cup \Sigma$ beschriftet sind, so dass
  - jeder innere Knoten mit einem Symbol aus N beschriftet ist,
  - jedes Blatt mit einem Symbol aus $N \cup \Sigma$ beschriftet ist und
  - wenn ein innerer Knoten v mit $A \in N$ beschriftet ist, so sind die Nachfolger von v (von links nach rechts) mit $w_1, \ldots, w_n$ beschriftet für eine Regel $A \to w_1 \cdots w_n \in P$.
- Ein *(vollständiger) Ableitungsbaum* ist part. Ableitungsbaum, so dass
  - die Wurzel mit S beschriftet ist,
  - jedes Blatt mit einem Symbol aus Σ beschriftet ist.

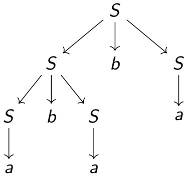
Ableitungsbäume Beispiel

- $P = \{S \longrightarrow SbS, S \longrightarrow a\}$
- Drei *Ableitungen* des Wortes *ababa*:
  1) $S \vdash SbS \vdash abS \vdash abSbS \vdash ababS \vdash ababa$
  2) $S \vdash SbS \vdash abS \vdash abSbS \vdash abSba \vdash ababa$
  3) $S \vdash SbS \vdash Sba \vdash SbSba \vdash Sbaba \vdash ababa$
- Die zugehörigen *Ableitungsbäume*:

für 1) und 2):



für 3):



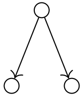
Ableitungsbäume

- Ein Ableitungsbaum kann für mehr als eine Ableitung stehen, und dasselbe Wort kann verschiedene Ableitungsbäume haben!
- Wenn eine Ableitung $A \vdash_G^* w_1 \dots w_n$ existiert, so existiert ein Ableitungsbaum mit Wurzel A und Blättern, die (von links nach rechts) mit $w_1, \dots, w_n$ beschriftet sind.
- Wenn ein Ableitungsbaum mit Wurzel A und Blättern, die (von links nach rechts) mit $w_1, \dots, w_n$ beschriftet sind existiert, so existiert eine Ableitung $A \vdash_G^* w_1 \dots w_n$.

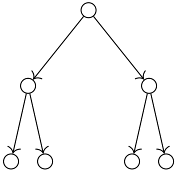
Beweis fortgesetzt

- Ein Baum der Tiefe $\leq t$ und der Verzweigungsgrad ≤ 2 hat maximal 2^t viele Blätter:

eine Ebene: ≤ 2 Blätter



zwei Ebenen: $\leq 2^2$ Blätter

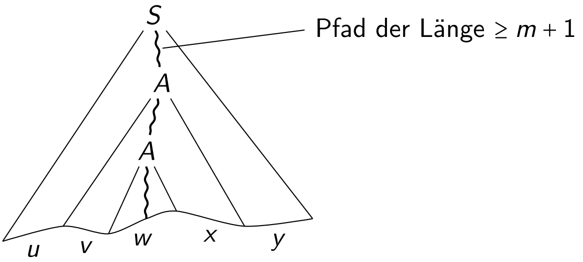


etc.

- Jeder Ableitungsbaum für z mit $|z| \geq n_0$ hat $|z| \geq n_0 = 2^{m+1}$ Blätter.
- Also gibt es einen Pfad (Weg von Wurzel zu Blatt) der Länge $\geq m + 1$ (Anzahl der Kanten), der also $> m + 1$ Knoten hat.

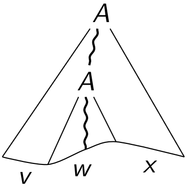
Beweis fortgesetzt

- Auf diesem Pfad kommen $> m+1$ Symbole vor, davon $> m$ Nichtterminale.
- Da es nur m verschiedene Nichtterminale gibt, kommt ein Nichtterminal A zweimal vor.
- Betrachte erste Wiederholung eines Nichtterminals A von den Blättern aus gesehen. Wir erhalten u, v, w, x, y:



Beweis fortgesetzt

- Dann der Teilbaum



die Tiefe $\leq m+1$, was $|vwx| \leq 2^{m+1} = n_0$ zeigt.

- Es gilt: $S \vdash_G^* uAy, \quad A \vdash_G^* vAx, \quad A \vdash_G^* w$, woraus folgt:

$$S \ \vdash_G^* \ uAy \ \vdash_G^* \ uv^iAx^iy \ \vdash_G^* \ uv^kwx^ky.$$

- Außerdem gilt $vx \neq \varepsilon$: Da G ε-frei ist, wäre sonst $A \vdash_G^+ vAx$ nur bei Anwesenheit von Regeln der Form $A \longrightarrow B$ möglich. □

Abschlusseigenschaften kontextfreier Sprachen

Abschlusseigenschaften kontextfreier Sprachen

Die Klasse $\mathcal{L}_2$ der kontextfreien Sprachen ist unter Vereinigung, Konkatenation und Kleene-Stern abgeschlossen.

Beweis. Seien $G_i = (N_i, \Sigma, P_i, S_i), i = 1, 2$ kontextfreie Grammatiken. O. B. d. A. gilt $N_1 \cap N_2 = \varnothing$.

1) $G := (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \longrightarrow S_1, S \longrightarrow S_2\}, S)$ mit $S \notin N_1 \cup N_2$ ist eine kontextfreie Grammatik mit $L(G) = L(G_1) \cup L(G_2)$.
2) $G' := (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \longrightarrow S_1 S_2\}, S)$ mit $S \notin N_1 \cup N_2$ ist eine kontextfreie Grammatik mit $L(G') = L(G_1) \cdot L(G_2)$.
3) $G'' := (N_1 \cup \{S\}, \Sigma, P_1 \cup \{S \longrightarrow \varepsilon, S \longrightarrow SS_1\}, S)$ mit $S \notin N_1$ ist eine kontextfreie Grammatik für $L(G_1)^*$.

Abschlusseigenschaften kontextfreier Sprachen

Satz

Die Klasse $\mathcal{L}_2$ der kontextfreien Sprachen ist *nicht* unter Schnitt und Komplement abgeschlossen.

Beweis

- Betrachte

$$L_1 = \{a^n b^n c^m | n \geq 1, m \geq 1\} \text{ und } L_2 = \{a^m b^n c^n | n \geq 1, m \geq 1\}.$$

- Die folgenden kontextfreien Grammatiken erzeugen L_1 bzw. L_2, also sind L_1 und L_2 kontextfrei.
    - $G_1 = (\{S, A, C\}, \{a, b, c\}, P, S)$ mit $P = \{S \to AC, A \to aAb \mid ab, C \to cC \mid c\}$.
    - $G_2 = (\{S, A, B\}, \{a, b, c\}, P, S)$ mit $P = \{S \to AB, A \to aA \mid a, B \to bBc \mid bc\}$.

Abschlusseigenschaften kontextfreier Sprachen

Die Sprachen

$$L_1 = \{a^n b^n c^m | n \geq 1, m \geq 1\} \text{ und } L_2 = \{a^m b^n c^n | n \geq 1, m \geq 1\}$$

sind in $\mathcal{L}_2$.

Schnitt:

Es gilt

$$\{a^n b^n c^m | n \geq 1, m \geq 1\} \cap \{a^m b^n c^n | n \geq 1, m \geq 1\} = \{a^n b^n c^n | n \geq 1\}.$$

Wir haben bereits gezeigt, dass $\{a^n b^n c^n | n \geq 1\} \notin \mathcal{L}_2$. Also ist $\mathcal{L}_2$ nicht unter $\cap$ abgeschlossen.

Komplement:

Es gilt $L_1 \cap L_1 = \overline{\overline{L_1} \cup \overline{L_2}}$.

Wäre $\mathcal{L}_2$ unter Komplement abgeschlossen, so wäre es auch unter Schnitt abgeschlossen. Also ist $\mathcal{L}_2$ nicht unter Komplement abgeschlossen. $\square$

Abschlusseigenschaften

Klasse	$\cap$	$\cup$	$\overline{}$	$\cdot$	$*$
$\mathcal{L}_0$	✓	✓	✗	✓	✓
$\mathcal{L}_1$	✓	✓	✓	✓	✓
$\mathcal{L}_2$	✗	✓	✗	✓	✓
$\mathcal{L}_3$	✓	✓	✓	✓	✓