LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Formale Sprachen und Komplexität
Sommersemester 2019

Zentralübung 06.06.2019: Pumping-Lemma für CFLs, Ogdens Lemma, CYK-Algorithmus

Prof. Dr. David Sabel

LFE Theoretische Informatik

Letzte Änderung der Folien: 6. Juni 2019

Wiederholung: Das Pumping-Lemma für CFLs

Lemma (Pumping-Lemma für CFLs)

Sei L eine kontextfreie Sprache. Dann gibt es eine Zahl $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$, das Mindestlänge n hat (d. h. $|z| \geq n$), als $z = uvwxy$ geschrieben werden kann, so dass gilt:

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$.

Beispiel

Aufgabe

Sei $L = \{w\overline{w} \mid w \in \{a,b\}^*\}$ wobei $\overline{w}$ das Wort w rückwärts gelesen ist.

1. Zeige, dass L kontextfrei ist.
2. Zeige, dass L die Pumping-Eigenschaft erfüllt.

1) Die CFG $G = (\{S\}, \{a,b\}, \{S \to aSa \mid bSb \mid \varepsilon\}, S)$ erzeugt L.

2) L erfüllt die Pumping-Eigenschaft:

- Wir wählen $n = 2$.
- Sei $z \in L$ mit $|z| \geq 2$. Dann ist $z = a_1 \cdots a_k a_k \cdots a_1$ mit $k \geq 1$.
- Zerlege $z = uvwxy$ mit
  $u = a_1 \cdots a_{k-1}, \quad v = a_k, \quad w = \varepsilon, \quad x = a_k, \quad y = a_{k-1} \cdots a_1$
- $|vwx| = 2 \geq n$ und $|vx| = 2 \geq 1$
- Dann gilt $uv^iwx^iy = a_1 \cdots a_{k-1} a_k^i a_k^i a_{k-1} \cdots a_1 \in L$ für alle $i \in \mathbb{N}$

Kontextfreiheit Widerlegen mit Pumping-Lemma als Spiel

Sei L die formale Sprache.

1. Der **Gegner** wählt die Zahl $n \in \mathbb{N}_{>0}$.
2. **Wir** wählen das Wort $z \in L$ mit $|z| \geq n$.
3. Der **Gegner** wählt die Zerlegung
   $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$
4. **Wir** gewinnen das Spiel, wenn wir ein $i \geq 0$ angeben können, sodass $uv^iwx^iw \notin L$.

Wenn wir **für jede Wahl des Gegners** das Spiel gewinnen können, dann haben wir gezeigt, dass L nicht kontextfrei ist.

Aufgabe

Aufgabe

Zeige, dass $L = \{a^m b a^m b a^m \mid m \in \mathbb{N}\}$ nicht kontextfrei ist.

Beweis mit Pumping-Lemma:

- Sei n beliebig. (vom Gegner gewählt).
- Wir wählen $z = a^n b a^n b a^n$
- Sei $z = uvwxy$ mit $|vwx| \leq n$ und $|vx| \geq 1$ (vom Gegner zerlegt)
- Dann kann vwx nicht zwei $b's$ enthalten.
- Fall: vx enthält ein b. Dann kann uv^0wx^0y nicht in L liegen, da das b entfernt wurde.
- Fall vx enthält kein b. Dann $uv^2wx^2y \notin L$, da maximal zwei a-Folgen aufgepumpt wurden, die dritte a-Folge aber noch aus n vielen a's besteht (und die Trennung durch b noch vorhanden ist).

Aufgabe

Aufgabe

Zeige, dass die Sprache
$L = \{a^i b^j c^j d^j \mid i \in \mathbb{N}_{>0}, j \in \mathbb{N}\} \cup (\{b\}^*\{c\}^*\{d\}^*)$ die Pumping-Eigenschaft für CFLs erfüllt.

- Wähle $n = 4$, sei $z \in L$ mit $|z| \geq 4$ mit $z = a_1 \cdots a_{|z|}$.
- Fall: $z \in (\{b\}^*\{c\}^*\{d\}^*)$
  Wähle die Zerlegung $u = v = w = \varepsilon$, $x = a_1$, $y = a_2 \cdots a_{|z|}$.
  Dann gilt $z = uvwxy$, $|vx| \geq 1, |vwx| \leq 4$ und
  $uv^iwx^iy = a_1^i y \in (\{b\}^*\{c\}^*\{d\}^*)$ für alle $i \in \mathbb{N}$.
- Fall: $z \in \{a^i b^j c^j d^j \mid i \in \mathbb{N}_{>0}, j \in \mathbb{N}\}$
  Wähle die Zerlegung $u = v = w = \varepsilon$, $x = a_1 = a$,
  $y = a_2 \cdots a_{|z|}$. Dann gilt $z = uvwxy$, $|vx| \geq 1, |vwx| \leq 4$ und
  $uv^iwx^iy = a^i y \in L$ für alle $i \in \mathbb{N}$.

Aber: L ist nicht kontextfrei (wird gleich gezeigt)!

Ogdens Lemma

- Verallgemeinerung des Pumping-Lemmas für CFLs: Das sogenannte Ogdens Lemma
- benannt nach William Ogden

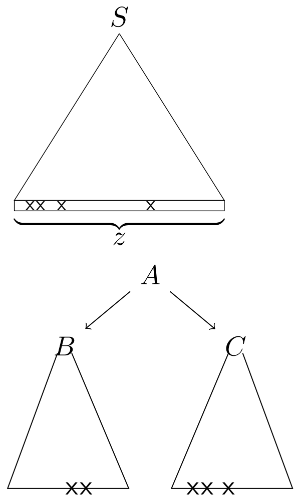
Ogdens Lemma

Sei L eine kontextfreie Sprache. Dann gibt es eine Zahl $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$, das Mindestlänge n hat und in dem mindestens n Zeichen markiert sind, als $z = uvwxy$ geschrieben werden kann, so dass gilt:

- vx enthält mindestens ein markiertes Zeichen
- vwx enthält höchstens n markierte Zeichen
- für alle $i \geq 0$: $uv^iwx^iy \in L$.

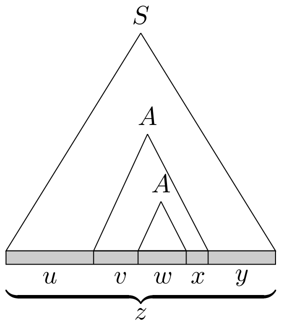
Ogdens Lemma: Beweis

- Sei L kontextfrei und $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-NF mit $L(G) = L \setminus \{\varepsilon\}$
- Sei $z \in L$ mit $|z| \geq 2^{|V|+1} = n$
- Seien $\geq n$ Zeichen in z markiert.
- Syntaxbaum ist bis auf letzte Schicht ($A \to a$) Binärbaum.
- Wähle Pfad im Syntaxbaum: Für Knoten mit A markiert, und Kindern B und C, wähle den Teilbaum, dessen zugehöriges Wort mehr markierte Zeichen enhält.
- Ein Knoten ist ein Verzweigungsknoten, wenn die zu beiden Kindern zugehörigen Worte noch markierte Zeichen enthalten



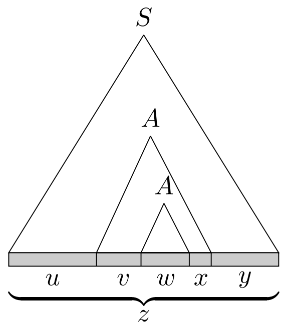
Ogdens Lemma: Beweis (2)

- Nach jedem Verzweigungsknoten ist die Menge der markierten Buchstaben höchstens halbiert
- Daher hat der gewählte Pfad $\geq |V| + 1$ Verzweigungsknoten.
- Betrachte Teilpfad von unten, der genau $|V| + 1$ Verzweigungsknoten enthält.
- Dort wird eine Variable $A \in V$ doppelt besucht.
- Das induziert Zerlegung $z = uvwxy$, wobei das obere A den Teilbaum T_{vwx} und das untere A den Teilbaum T_w erzeugt.
- T_w kann nur im linken oder im rechten Unterbaum von T_{vwx} liegen, aber beide Unterbäume enthalten noch Markierungen (da A ein Verzweigungsknoten ist)
- Daher enthält vx mindestens ein markiertes Zeichen



Ogdens Lemma: Beweis (3)

- Der Pfad vom oberen A aus enthält $|V| + 1$ Verzweigungsknoten und kann daher maximal $2^{|V|+1} = n$ Zeichen markieren.
- Daher enthält vwx maximal n markierte Zeichen.
- $uv^iwx^iy \in L(G)$ da $S \Rightarrow_G^* uAv$, $A \Rightarrow_G^* vAx$ und $A \Rightarrow_G^* w$.



Pumping-Lemma ist Spezialfall von Ogdens Lemma

Satz

Das Pumping-Lemma für CFLs ist ein Spezialfall von Ogdens Lemma.

Beweis:
Markiere alle Zeichen in z (anstelle von mindestens n), dann erhält man das Pumping-Lemma für CFLs.

Aufgabe

Aufgabe

Zeige, dass die Sprache
$L = \{a^ib^jc^jd^j \mid i \in \mathbb{N}_{>0}, j \in \mathbb{N}\} \cup (\{b\}^*\{c\}^*\{d\}^*)$ nicht kontextfrei ist, durch Verwendung von Ogdens Lemma.

Beweis:

- Sei $n \in \mathbb{N}_{>0}$. Wir wählen $z = ab^nc^nd^n \in L$ und markieren das Teilwort bc^nd.
- Sei $z = uvwxy$, wobei vx mindestens eines und vwx höchstens n markierte Zeichen enhält.
- Teilwort vwx enthält nicht b's, c's und d's
- Daher $uv^2wx^2y \notin L$ da dort die Anzahl an b's, c's und d's nicht gleich sein kann, und das Wort mit a beginnt.

Algorithmus 8: CYK-Algorithmus

Eingabe: CFG $G = (V, \Sigma, P, S)$ in Chomsky-NF und Wort $w = a_1 \cdots a_n \in \Sigma^*$
Ausgabe: `Ja`, wenn $w \in L(G)$ und `Nein`, wenn $w \notin L(G)$
Beginn

- **für** $i = 1$ *bis* n **tue**
  - $V(i,1) = \{A \in V \mid A \to a_i \in P\}$
- **für** $j = 2$ *bis* n **tue**
  - **für** $i = 1$ *bis* $n + 1 - j$ **tue**
    - $V(i,j) = \emptyset$;
    - **für** $k = 1$ *bis* $j - 1$ **tue**
      - $$V(i,j) = V(i,j) \cup \left\{ A \in V \;\middle|\; \begin{array}{l} A \to BC \in P, \\ B \in V(i,k), \\ C \in V(i+k, j-k) \end{array} \right\}$$
- **wenn** $S \in V(1,n)$ **dann**
  - return `Ja`
- **sonst**
  - return `Nein`

Aufgabe

Sei $G = (\{S, A, B, C, D, E\}, \{a, b\}, P, S)$ wobei P :

$$\begin{array}{lll} S \to SE \mid AB \mid CE, & A \to a, & B \to b \\ C \to AB \mid DD, & D \to AB \mid ED, & E \to AB \end{array}$$

Verwende den CYK-Algorithmus, um zu entscheiden, ob $ababab \in L$ gilt

	a	b	a	b	a	b
j \ i	1	2	3	4	5	6
1	A	B	A	B	A	B
2	S, E, C, D	$\emptyset$	S, E, C, D	$\emptyset$	S, E, C, D	
3	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$		
4	S, C, D	$\emptyset$	S, C, D			
5	$\emptyset$	$\emptyset$				
6	D, S, C					

Da $S \in V(1,6)$ ist $ababab \in L$