Theoretische Grundlagen der Informatik

Vorlesung am 18.01.2011

INSTITUT FÜR THEORETISCHE INFORMATIK

Ergebnisse zum Wortproblem

Karlsruher Institut für Technologie

- **Typ-0 Grammatik.** Das Wortproblem ist nicht entscheidbar.
- **Typ-1 Grammatik.** Das Wortproblem ist NP-vollständig.
- **Typ-2 Grammatik.** Das Wortproblem ist in polynomieller Zeit lösbar.
- **Typ-3 Grammatik.** Das Wortproblem ist in linearer Zeit lösbar.

Das Pumping-Lemma für kontextfreie Sprachen

Pumping-Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L
gibt es eine Konstante $n \in \mathbb{N}$,
so dass sich jedes Wort $z \in L$ mit $|z| \geq n$

so als

- $z = uvwxy$

schreiben lässt, dass

- $|vx| \geq 1$,
- $|vwx| \leq n$ und
- für alle $i \geq 0$ das Wort $uv^iwx^iy \in L$ ist.

Ogden's Lemma für kontextfreie Sprachen

Ogden's Lemma für kontextfreie Sprachen

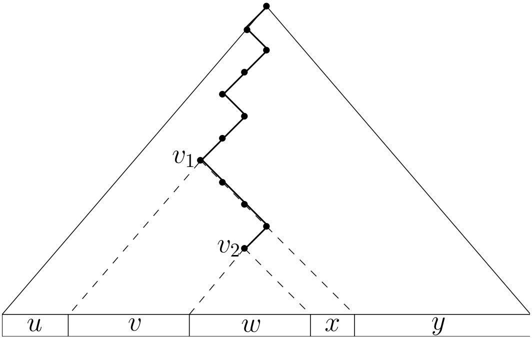
Für jede kontextfreie Sprache L
gibt es eine Konstante $n \in \mathbb{N}$,
so dass für jedes Wort $z \in L$ mit $|z| \geq n$ gilt:

Wenn wir in z mindestens n Buchstaben markieren, so lässt sich z so als $z = uvwxy$ schreiben,

- dass von den mindestens n markierten Buchstaben
  - mindestens einer zu vx gehört und
  - höchstens n zu vwx gehören und
- für alle $i \geq 0$ das Wort $uv^iwx^iy \in L$ ist.

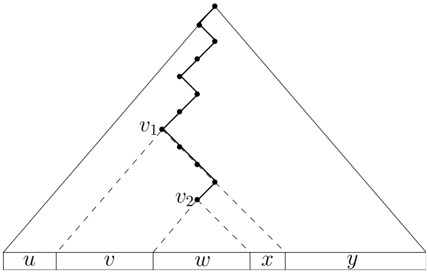
Beweis

- Sei L kontextfreie Sprache
- Sei G Grammatik zu L mit Variablen V in Chomsky-Normalform, d.h. alle Regeln sind von der Form $A \to BC$ oder $A \to a$.
- Setze $n := 2^{|V|+1}$.
- Wähle beliebiges Wort $z \in L$ mit $|z| \geq n$
- Betrachte einen Syntaxbaum T zu z.



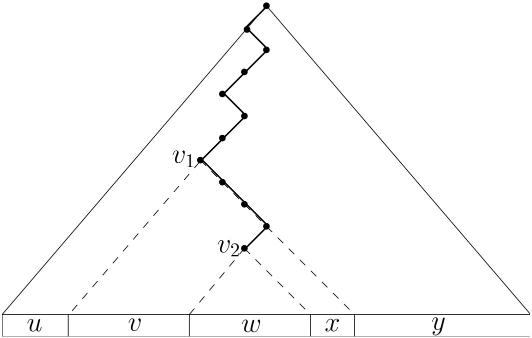
Beweis

- T hat $|z|$ Blätter und alle inneren Knoten außer den Vorgängern der Blätter haben Grad 2, ansonsten Grad 1.
- Seien mindestens n Blätter markiert.
- Durchlaufe einen Weg von der Wurzel zu einem Blatt.
  Wähle stets den Nachfolger, auf dessen Seite die größere Anzahl markierter Blätter liegt.
- Nenne Knoten auf dem Weg, für die rechter und linker Unterbaum markierte Blätter hat, **Verzweigungsknoten**.



Beweis

- Wegen $n > 2^{|V|}$ liegen auf dem Weg mindestens $|V| + 1$ Verzweigungsknoten
- Von den letzten $|V| + 1$ Verzweigungsknoten entsprechen mindestens zwei Knoten v_1, v_2 derselben Variablen A.
- Sei vwx Wort unter Teilbaum mit Wurzel v_1
- Sei w Wort unter Teilbaum mit Wurzel v_2.
- Damit sind u und y eindeutig bestimmt.



Beweis

- Da v_1 Verzweigungsknoten ist, enthält vx mindestens einen markierten Buchstaben.
- Da der Unterbaum von v_1 inkl. v_1 nur $|V| + 1$ Verzweigungsknoten enthält, gibt es in vwx höchstens $2^{|V|+1} = n$ markierte Buchstaben.
- Zu G existieren die Ableitungen

$$S \xrightarrow{*} uAy, \qquad A \xrightarrow{*} vAx, \qquad A \xrightarrow{*} w.$$

Daraus kann z abgeleitet werden durch

$$S \xrightarrow{*} uAy \xrightarrow{*} uvAxy \xrightarrow{*} uvwxy = z,$$

aber auch uv^iwx^iy für jedes $i \geq 1$ durch

$$S \xrightarrow{*} uAy \xrightarrow{*} uvAxy \xrightarrow{*} uv^2Ax^2y \xrightarrow{*} \cdots \rightarrow uv^iAx^iy \rightarrow uv^iwx^iy.$$

Also ist auch $uv^iwx^iy \in L$ für $i \geq 0$.

Bemerkung

- Der Spezialfall von Odgen's Lemma, in dem alle Buchstaben von z markiert sind, ist gerade das Pumping–Lemma.

Satz:
Die Chomsky-Hierarchie ist echt, d.h.

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0 \ ,$$

wobei $\mathcal{L}_i, 0 \leq i \leq 3$, Klasse der durch Typ-$i$-Grammatiken erzeugten Sprachen.

Beweis - Teil 1

Es gibt eine kontextfreie Sprache, die nicht regulär ist.

Die Sprache

$$L = \{a^i b^i | i \geq 1\}$$

ist kontextfrei und wird durch die Grammatik

$$\begin{aligned} V &= \{S\} \\ \Sigma &= \{0, 1\} \\ R &= \{S \to 01 \mid 0S1\} \,. \end{aligned}$$

erzeugt. Sie ist aber nicht regulär.
(Beispiel (2) zum Pumping-Lemma, Vorlesung vom 26.1.2010)

Beweis - Teil 2

Es gibt eine kontextsensitive Sprache, die nicht kontextfrei ist.

Die Sprache

$$L = \{a^i b^i c^i | i \geq 1\}$$

ist kontextsensitiv.

Beweis

- L kontextsensitiv $\Leftrightarrow$ es gibt NTM mit linearem Speicherbedarf für L
- Eingabe $w \in \{a, b, c\}^*$
- Überprüfe deterministisch, ob $w = a^i b^j c^k$
- Überprüfe deterministisch, ob $j = i$ und $k = i$
- Speicherbedarf: $i + j + k$, also linear
- $\Rightarrow L$ kontextsensitiv

Beweis - Teil 2

Es gibt eine kontextsensitive Sprache, die nicht kontextfrei ist.

Die Sprache

$$L = \{a^i b^i c^i | i \geq 1\}$$

ist nicht kontextfrei.

Beweis - Teil 2

Pumping-Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L
gibt es eine Konstante $n \in \mathbb{N}$,
so dass sich jedes Wort $z \in L$ mit $|z| \geq n$

so als

- $z = uvwxy$

schreiben lässt, dass

- $|vx| \geq 1$,
- $|vwx| \leq n$ und
- für alle $i \geq 0$ das Wort $uv^iwx^iy \in L$ ist.

Beweis - Teil 2

Die Sprache $L = \{a^i b^i c^i | i \geq 1\}$ ist nicht kontextfrei.

Beweis

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung $z = uvwxy$ wie im PL gefordert:
  - $|vx| \geq 1$,
  - $|vwx| \leq n$ und
  - für alle $i \geq 0$ ist das Wort $uv^i wx^i y \in L$.

Beweis - Teil 2

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

Beweis

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung $z = uvwxy$ wie im PL gefordert:
  - $|vx| \geq 1$,
  - $|vwx| \leq n$ und
  - für alle $i \geq 0$ ist das Wort $uv^i wx^i y \in L$.

- Fallunterscheidung, Fall 1: vwx besteht nur aus a und b
  - Dann enthalt vx mindestens ein a oder b.
  - Damit ist $uv^0 wx^0 y = a^i b^j c^n \notin L$ weil entweder $i < n$ oder $j < n$.
  - Dies ist ein Widerspruch zum PL.

Beweis - Teil 2

Die Sprache $L = \{a^i b^i c^i | i \geq 1\}$ ist nicht kontextfrei.

Beweis

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung $z = uvwxy$ wie im PL gefordert:
  - $|vx| \geq 1$,
  - $|vwx| \leq n$ und
  - für alle $i \geq 0$ ist das Wort $uv^i wx^i y \in L$.
- Fallunterscheidung, Fall 2: vwx besteht nur aus b und c
  - Dann enthalt vx mindestens ein b oder c.
  - Damit ist $uv^0 wx^0 y = a^n b^i c^j \notin L$ weil entweder $i < n$ oder $j < n$.
  - Dies ist ein Widerspruch zum PL.

Beweis - Teil 2

Die Sprache $L = \{a^i b^i c^i | i \geq 1\}$ ist nicht kontextfrei.

Ogden's Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L
gibt es eine Konstante $n \in \mathbb{N}$,
so dass für jedes Wort $z \in L$ mit $|z| \geq n$ gilt:

Wenn wir in z mindestens n Buchstaben markieren, so lässt sich z so als $z = uvwxy$ schreiben,

- dass von den mindestens n markierten Buchstaben
  - mindestens einer zu vx gehört und
  - höchstens n zu vwx gehören und
- für alle $i \geq 0$ das Wort $uv^iwx^iy \in L$ ist.

Beweis - Teil 2

Die Sprache $L = \{a^i b^i c^i | i \geq 1\}$ ist nicht kontextfrei.

Alternativer Beweis mit Odgen's Lemma

- Annahme: L sei kontextfrei.
- Sei dann n wie in Odgen's Lemma gefordert.
- Wähle das Wort $z = a^{n+1} b^{n+1} c^{n+1} \in L$.
- Markiere alle b.
- Damit enthält vwx mindestens ein b aber kein a oder kein c.
- Es enthalte vwx kein c (anderer Fall analog)
- Damit ist $uv^0 wx^0 y = a^i b^j c^n \notin L$ weil entweder $i < n$ oder $j < n$.
- Dies ist ein Widerspruch zu Odgen's Lemma.

Beweis - Teil 3

Es gibt eine semi-entscheidbare Sprache, die nicht kontextsensitiv ist.

Es sei L_u die universelle Sprache.

Wiederholung

Die **universelle Sprache** L_u über $\{0, 1\}$ ist definiert durch

$$L_u := \{wv \mid v \in L(T_w)\} .$$

L_u ist also die Menge aller Wörter wv für die T_w bei der Eingabe v hält und v akzeptiert.

Beweis - Teil 3

Es gibt eine semi-entscheidbare Sprache, die nicht kontextsensitiv ist.

Es sei L_u die universelle Sprache.

- L_u ist semi-entscheidbar, aber nicht entscheidbar (Kapitel 3).
- Wegen der Semi-entscheidbarkeit gilt $L_u \in \mathcal{L}_0$.

- Annahme: $L_u \in \mathcal{L}_1$.
- Dann gibt es eine NTM, die L_u mit linearem Speicher erkennt.
- Mit linearem Speicher können nur exponentiell viele verschiedene Konfigurationen auftreten.
- Diese könnte durch eine DTM durch Ausprobieren aller möglichen Konfigurationen simuliert werden.
- Dies wäre ein Widerspruch zur Nichtentscheidbarkeit von L_u.