TECHNISCHE UNIVERSITÄT DRESDEN

FORMALE SYSTEME

14. Vorlesung: Abschlusseigenschaften kontextfreier Sprachen

Markus Krötzsch
Lehrstuhl Wissensbasierte Systeme

TU Dresden, 27. November 2017

Rückblick: Das Pumping Lemma

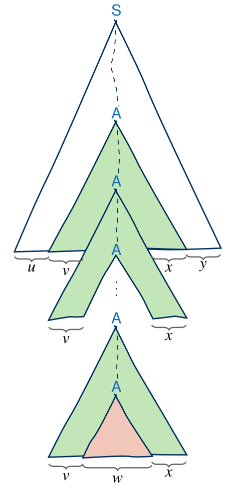
Pumpen für kontextfreie Sprachen

Satz (Pumping Lemma): Für jede kontextfreie Sprache **L**
gibt es eine Zahl $n \geq 0$, so dass gilt:
für jedes Wort $z \in \mathbf{L}$ mit $|z| \geq n$
gibt es eine Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$, s.d.:
für jede Zahl $k \geq 0$ gilt: $uv^kwx^ky \in \mathbf{L}$

Beispiel: Für die Sprache $\{\mathtt{a}^i\mathtt{b}^i \mid i \geq 0\}$ gilt der Satz. Wir wählen $n = 2$. Sei $z = \mathtt{a}^i\mathtt{b}^i$ mit $i \geq 1$ ein beliebiges Wort mit $|z| \geq 2$. Wir wählen die Zerlegung $u = \mathtt{a}^{i-1}$, $v = \mathtt{a}$, $w = \epsilon$, $x = \mathtt{b}$ und $y = \mathtt{b}^{i-1}$.
Dann ist $uv^kwx^ky = \mathtt{a}^{i-1}\mathtt{a}^k\mathtt{b}^k\mathtt{b}^{i-1} = \mathtt{a}^{i+k-1}\mathtt{b}^{i+k-1} \in \mathbf{L}$ für alle $k \geq 0$.

Ableitungsbäume aufpumpen

Die Idee des Lemmas lässt sich gut am Ableitungsbaum darstellen:



Abgeleitetes Wort:
uv^kwx^ky

Beispiel

CNF-Grammatik für $\{\mathtt{a}^i\mathtt{b}^i \mid i \geq 1\}$:

S → AB | AC

A → a

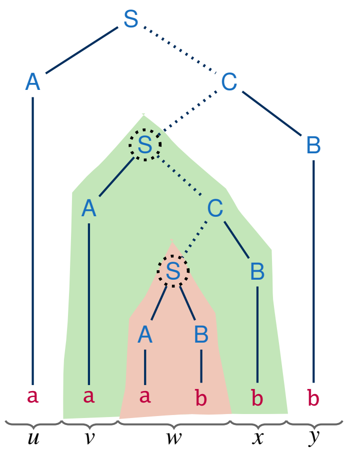
B → b

C → SB

Konstante aus dem Beweis des Pumping-Lemma: $n = 2^4 = 16$

Aber man kann auch schon ab $n = 4$ pumpen

Beispiel: Ableitung für aaabbb



Beispiel (2)

CNF-Grammatik für $\{\mathtt{a}^i\mathtt{b}^i \mid i \geq 1\}$:

S → AB | AC

A → a

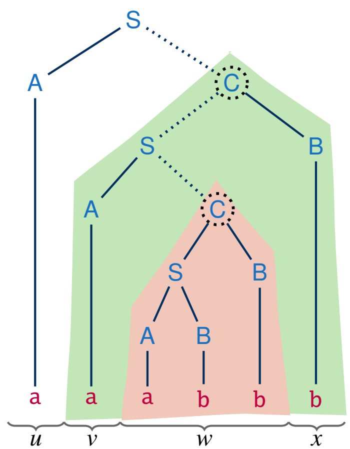
B → b

C → SB

Konstante aus dem Beweis des Pumping-Lemma: $n = 2^4 = 16$

Aber man kann auch schon ab $n = 4$ pumpen

Beispiel: Ableitung für aaabbb



Beispiel (3)

CNF-Grammatik für $\mathtt{ab}^+\mathtt{a}$:

S → AB

A → a

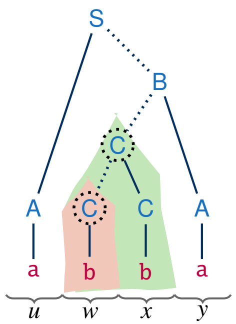
B → CA

C → CC | b

Konstante aus dem Beweis des Pumping-Lemma: $n = 2^4 = 16$

Aber man kann auch schon ab $n = 4$ pumpen

Beispiel: Ableitung für abba



$v = \epsilon$

Abschlusseigenschaften kontextfreier Sprachen

Rückblick: Reguläre Sprachen

Satz: Wenn $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$ reguläre Sprachen sind, dann beschreiben auch die folgenden Ausdrücke reguläre Sprachen:

(1) $\mathbf{L_1} \cup \mathbf{L_2}$ (Abschluss unter Vereinigung)
(2) $\mathbf{L_1} \cap \mathbf{L_2}$ (Abschluss unter Schnitt)
(3) $\overline{\mathbf{L}}$ (Abschluss unter Komplement)
(4) $\mathbf{L_1} \circ \mathbf{L_2}$ (Abschluss unter Konkatenation)
(5) $\mathbf{L}^*$ (Abschluss unter Kleene-Stern)

Wie sieht es bei den kontextfreien Sprachen aus?

Abschluss für kontextfreie Sprachen?

Bei kontextfreien Sprachen ergibt sich ein anderes Bild:

Satz: Wenn $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$ kontextfreie Sprachen sind, dann beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

(1) $\mathbf{L_1} \cup \mathbf{L_2}$ (Abschluss unter Vereinigung)
(2) $\mathbf{L_1} \circ \mathbf{L_2}$ (Abschluss unter Konkatenation)
(3) $\mathbf{L}^*$ (Abschluss unter Kleene-Stern)

Aber:

Satz: Es gibt kontextfreie Sprachen $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$, so dass die folgenden Ausdrücke keine kontextfreien Sprachen sind:

(1) $\mathbf{L_1} \cap \mathbf{L_2}$ (Nichtabschluss unter Schnitt)
(2) $\overline{\mathbf{L}}$ (Nichtabschluss unter Komplement)

Abschlusseigenschaften von Typ-2-Sprachen

Satz: Wenn $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$ kontextfreie Sprachen sind, dann beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

(1) $\mathbf{L_1} \cup \mathbf{L_2}$ Abschluss unter Vereinigung
(2) $\mathbf{L_1} \circ \mathbf{L_2}$ (Abschluss unter Konkatenation)
(3) $\mathbf{L}^*$ (Abschluss unter Kleene-Stern)

Abschluss unter Vereinigung

Wir konstruieren eine Vereingungsgrammatik

Gegeben seien zwei formale Grammatiken $G_1 = \langle V_1, \Sigma, P_1, \mathsf{S}_1 \rangle$ und $G_2 = \langle V_2, \Sigma, P_2, \mathsf{S}_2 \rangle$ mit $V_1 \cap V_2 = \emptyset$ (o.B.d.A.).

Die Vereinigungsgrammatik $G_1 \uplus G_2$ ist gegeben durch

$$G_1 \uplus G_2 = \langle V_1 \cup V_2 \cup \{\mathsf{S}\}, \Sigma, P_1 \cup P_2 \cup \{\mathsf{S} \to \mathsf{S}_1 \mid \mathsf{S}_2\}, \mathsf{S} \rangle,$$

wobei S ein neues Startsymbol ist, das nicht in $V_1 \cup V_2$ vorkommt.

In Worten: Die neuen Ableitungsregeln $\mathsf{S} \to \mathsf{S}_1 \mid \mathsf{S}_2$ ermöglichen es, dass $G_1 \uplus G_2$ entweder Wörter aus G_1 oder aus G_2 generiert.

Es ist daher leicht zu sehen:

Satz: $\mathbf{L}(G_1 \uplus G_2) = \mathbf{L}(G_1) \cup \mathbf{L}(G_2)$.

Vereinigungen in Typ 2?

Bisher haben wir nur erkannt:

Satz: $\mathbf{L}(G_1 \uplus G_2) = \mathbf{L}(G_1) \cup \mathbf{L}(G_2)$.

Für die Abschlusseigenschaft sollte noch mehr gelten:

Satz: Wenn G_1 und G_2 kontextfrei sind, dann ist auch $G_1 \uplus G_2$ kontextfrei.

Das ist leicht zu sehen, da die zusätzliche Regel kontextfrei ist.

Daher gilt sogar noch ein stärkerer Satz:

Satz: Wenn G_1 und G_2 von Typ $i \in \{2, 1, 0\}$ sind, dann ist auch $G_1 \uplus G_2$ von Typ i.

$\leadsto$ Typ-0-Sprachen, kontextsensitive Sprachen und kontextfreie Sprachen sind unter Vereinigung abgeschlossen

(reguläre Sprachen auch, aber das haben wir anders gezeigt)

Abschlusseigenschaften von Typ-2-Sprachen

Satz: Wenn $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$ kontextfreie Sprachen sind, dann beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

(1) $\mathbf{L_1} \cup \mathbf{L_2}$ (Abschluss unter Vereinigung)

(2) $\mathbf{L_1} \circ \mathbf{L_2}$ Abschluss unter Konkatenation

(3) $\mathbf{L}^*$ (Abschluss unter Kleene-Stern)

Konkatenation von Grammatiken

Wir erinnern uns: $\mathbf{L_1} \circ \mathbf{L_2} = \{w_1 w_2 \mid w_1 \in \mathbf{L_1} \text{ und } w_2 \in \mathbf{L_2}\}$

Es ist nicht schwer, eine passende Grammatik zu finden:

Gegeben seien zwei formale Grammatiken $G_1 = \langle V_1, \Sigma, P_1, \mathsf{S}_1 \rangle$ und $G_2 = \langle V_2, \Sigma, P_2, \mathsf{S}_2 \rangle$ mit $V_1 \cap V_2 = \emptyset$ (o.B.d.A.).

Die Grammatik $G_1 \circ G_2$ ist gegeben durch

$$G_1 \circ G_2 = \langle V_1 \cup V_2 \cup \{\mathsf{S}\}, \Sigma, P_1 \cup P_2 \cup \{\mathsf{S} \to \mathsf{S}_1\mathsf{S}_2\}, \mathsf{S} \rangle,$$

wobei S ein neues Startsymbol ist, das nicht in $V_1 \cup V_2$ vorkommt.

In Worten: Die neue Ableitungsregel $\mathsf{S} \to \mathsf{S}_1\mathsf{S}_2$ ermöglicht es, dass $G_1 \circ G_2$ Wörter aus G_1 gefolgt von Wörtern aus G_2 generiert.

Beispiel Konkatenation

Wir betrachten die Grammatiken

$$G_1: \quad \mathsf{S}_1 \to \mathsf{a}\mathsf{S}_1\mathsf{b} \mid \epsilon$$
$$G_2: \quad \mathsf{S}_2 \to \mathsf{c}\mathsf{S}_2 \mid \epsilon$$

Damit ist $\mathbf{L}(G_1) = \{\mathsf{a}^i\mathsf{b}^i \mid i \geq 0\}$ und $\mathbf{L}(G_2) = \{\mathsf{c}^k \mid k \geq 0\}$.

Die Konkatenation der Grammatiken enthält die folgenden Regeln:

$$G_1 \circ G_2: \qquad \mathsf{S} \to \mathsf{S}_1\mathsf{S}_2 \qquad \mathsf{S}_1 \to \mathsf{a}\mathsf{S}_1\mathsf{b} \mid \epsilon \qquad \mathsf{S}_2 \to \mathsf{c}\mathsf{S}_2 \mid \epsilon$$

Damit ergibt sich $\mathbf{L}(G_1 \circ G_2) = \{\mathsf{a}^i\mathsf{b}^i\mathsf{c}^k \mid i \geq 0, k \geq 0\}$.

Korrektheit

Hypothese: $\mathbf{L}(G_1 \circ G_2) = \mathbf{L}(G_1) \circ \mathbf{L}(G_2)$

Leider stimmt das nicht:

Gegenbeispiel: G_1 sei die Grammatik mit der einen Regel $\mathsf{S}_1 \to \mathsf{a}$ und G_2 die Grammatik mit den Regeln $\mathsf{S}_2 \to \mathsf{b}$ und $\mathsf{aS}_2 \to \mathsf{c}$.

Dann ist $\mathbf{L}(G_1) = \{\mathsf{a}\}$ und $\mathbf{L}(G_2) = \{\mathsf{b}\}$.
Trotzdem erlaubt $G_1 \circ G_2$ die Ableitung $\mathsf{S} \Rightarrow \mathsf{S}_1\mathsf{S}_2 \Rightarrow \mathsf{aS}_2 \Rightarrow \mathsf{c}$.

Korrekt ist dagegen:

Satz: Wenn G_1 und G_2 kontextfrei sind, dann ist $\mathbf{L}(G_1 \circ G_2) = \mathbf{L}(G_1) \circ \mathbf{L}(G_2)$.
Zudem ist $G_1 \circ G_2$ in diesem Fall kontextfrei.

Beweis: siehe Tafel.

Damit erhalten wir den gewünschten Abschluss.

Abschlusseigenschaften von Typ-2-Sprachen

Satz: Wenn $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$ kontextfreie Sprachen sind, dann beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

(1) $\mathbf{L_1} \cup \mathbf{L_2}$ (Abschluss unter Vereinigung)

(2) $\mathbf{L_1} \circ \mathbf{L_2}$ (Abschluss unter Konkatenation)

(3) $\mathbf{L}^*$ Abschluss unter Kleene-Stern

Kleene-Stern für Grammatiken

Wir erinnern uns: $\mathbf{L}^* = \{w_1w_2\cdots w_i \mid i \geq 0, w_1, \ldots, w_i \in \mathbf{L}\} = \bigcup_{i\geq 0} \mathbf{L}^i$

Auch hier kann man leicht eine passende Grammatik finden:

Gegeben sei eine formale Grammatik $G = \langle V, \Sigma, P, \mathsf{S}\rangle$.

Die Grammatik G^* ist gegeben durch

$$G^* = \langle V \cup \{\mathsf{S}'\}, \Sigma, P \cup \{\mathsf{S}' \to \epsilon \mid \mathsf{SS}'\}, \mathsf{S}'\rangle,$$

wobei S' ein neues Startsymbol ist, das nicht in V vorkommt.

In Worten: Die neuen Ableitungsregeln $\mathsf{S}' \to \epsilon \mid \mathsf{SS}'$ ermöglichen es, dass G^* beliebig lange Ketten aus Wörtern aus G generiert.

Beispiel Kleene-Stern

Wir betrachten die Grammatik

$$G: \quad \mathsf{S} \to \mathsf{SaSb} \mid \mathsf{SbSa} \mid \epsilon$$

Übung: Welche Sprache erzeugt diese Grammatik?

Der Kleene-Abschluss dieser Grammatik ist

$$G^*: \quad \mathsf{S}' \to \epsilon \mid \mathsf{SS}' \qquad \mathsf{S} \to \mathsf{SaSb} \mid \mathsf{SbSa} \mid \epsilon$$

Übung: Welche Sprache erzeugt diese Grammatik?

Korrektheit beim Kleene-Abschluss

Wie schon beim Konkatenation funktioniert diese Operation auf kontextfreien Grammatiken wie erwünscht:

Satz: Wenn G kontextfrei ist, dann ist $\mathbf{L}(G^*) = \mathbf{L}(G)^*$.
Zudem ist G^* in diesem Fall kontextfrei.

Beweis: siehe Tafel.

Für nicht kontextfreie Grammatiken gilt dieser Satz im Allgemeinen nicht:

Beispiel: Die (kontextsensitive) Grammatik G mit der einen Regel $\mathsf{SS} \to \mathtt{0}$ kann nichts hervorbringen.
Trotzdem erlaubt die Grammatik G^* die Ableitung eines Wortes:

$$\mathsf{S}' \Rightarrow \mathsf{SS} \Rightarrow \mathtt{0}$$

Nicht-Abschlusseigenschaften

Satz: Es gibt kontextfreie Sprachen $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$, so dass die folgenden Ausdrücke keine kontextfreien Sprachen sind:

(1) $\mathbf{L_1} \cap \mathbf{L_2}$ Nichtabschluss unter Schnitt

(2) $\overline{\mathbf{L}}$ (Nichtabschluss unter Komplement)

Nicht-Abschluss unter ∩

Beweisansatz: Wir müssen kontextfreie Sprachen $\mathbf{L_1}$ und $\mathbf{L_2}$ finden, für die $\mathbf{L_1} \cap \mathbf{L_2}$ nicht kontextfrei ist.

Welche nicht-kontextfreien Sprachen kennen wir?

$$\{\mathtt{a}^i\mathtt{b}^i\mathtt{c}^i \mid i \geq 0\}$$

Welche ähnlichen kontextfreien Sprachen kennen wir?

$$\{\mathtt{a}^i\mathtt{b}^i\mathtt{c}^k \mid i \geq 0, k \geq 0\}$$

Nicht-Abschluss unter ∩ (2)

Satz: Es gibt kontextfreie Sprachen $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$, so dass $\mathbf{L_1} \cap \mathbf{L_2}$ keine kontextfreie Sprache ist.

Beweis: Die folgenden Sprachen sind kontextfrei:

$$\begin{array}{rr}
\{\mathtt{a}^i\mathtt{b}^i \mid i \geq 0\} & \text{(zuvor gezeigt)} \\
\{\mathtt{b}^i\mathtt{c}^i \mid i \geq 0\} & \text{(analog)} \\
\{\mathtt{a}^i \mid i \geq 0\} & \text{(regulär)} \\
\{\mathtt{c}^i \mid i \geq 0\} & \text{(regulär, zuvor gezeigt)}
\end{array}$$

Dank Abschluss unter Konkatenation sind also auch kontextfrei:

$$\mathbf{L_1} = \{\mathtt{a}^i\mathtt{b}^i \mid i \geq 0\} \circ \{\mathtt{c}^i \mid i \geq 0\} = \{\mathtt{a}^i\mathtt{b}^i\mathtt{c}^k \mid i \geq 0, k \geq 0\}$$
$$\mathbf{L_2} = \{\mathtt{a}^i \mid i \geq 0\} \circ \{\mathtt{b}^i\mathtt{c}^i \mid i \geq 0\} = \{\mathtt{a}^i\mathtt{b}^k\mathtt{c}^k \mid i \geq 0, k \geq 0\}.$$

Der Schnitt $\mathbf{L_1} \cap \mathbf{L_2}$ ist aber $\{\mathtt{a}^i\mathtt{b}^i\mathtt{c}^i \mid i \geq 0\}$ und also nicht kontextfrei (zuvor gezeigt). □

Nicht-Abschlusseigenschaften

Satz: Es gibt kontextfreie Sprachen $\mathbf{L}$, $\mathbf{L_1}$ und $\mathbf{L_2}$, so dass die folgenden Ausdrücke keine kontextfreien Sprachen sind:

(1) $\mathbf{L_1} \cap \mathbf{L_2}$ (Nichtabschluss unter Schnitt)

(2) $\overline{\mathbf{L}}$ Nichtabschluss unter Komplement

Nicht-Abschluss unter Komplement

Satz: Es gibt eine kontextfreie Sprache $\mathbf{L}$, so dass $\overline{\mathbf{L}}$ keine kontextfreie Sprache ist.

Beweis: Die Behauptung folgt unmittelbar aus dem bereits gezeigten:

- Angenommen, Typ-2-Sprachen wären unter Komplement abgeschlossen
- Wir wissen bereits, dass Typ-2-Sprachen unter Vereinigung abgeschlossen sind
- Laut den Gesetzen der Mengenlehre gilt:

$$\mathbf{L_1} \cap \mathbf{L_2} = \overline{\overline{\mathbf{L_1}} \cup \overline{\mathbf{L_2}}} \qquad \text{(De Morgan)}$$

- Daraus folgt, dass Typ-2-Sprachen unter Schnitt abgeschlossen sind – Widerspruch □

Eine Warnung zum Nicht-Abschluss

Nicht-Abschluss unter Schnitt und Komplement bedeutet nicht, dass Schnitte bzw. Komplemente kontextfreier Sprachen grundsätzlich nicht kontextfrei sein können.

Beispiel: Alle regulären Sprachen sind kontextfrei, aber ihre Schnitte und Komplemente sind weiterhin regulär, also auch kontextfrei.

Beispiel: Das Komplement der nicht-regulären Sprache $\mathbf{L} = \{a^i b^i \mid i \geq 0\}$ über dem Alphabet $\{a, b\}$ ist die Vereinigung der folgenden Sprachen:

$$\mathbf{L_1} = \{a, b\}^*\{ba\}\{a, b\}^* \qquad \mathbf{L_2} = \{a\}^+\mathbf{L} \qquad \mathbf{L_3} = \mathbf{L}\{b\}^+$$

$\mathbf{L_1}$ ist regulär, also kontextfrei. $\mathbf{L_2}$ und $\mathbf{L_3}$ sind kontextfrei, da sie als Konkatenation zweier kontextfreier Sprachen entstehen. Die Vereinigung $\mathbf{L_1} \cup \mathbf{L_2} \cup \mathbf{L_3}$ ist also auch kontextfrei.

Ein nicht-kontextfreies Komplement

Unser Widerspruchsbeweis zum Abschluss unter Komplement liefert uns noch kein konkretes Beispiel eines nicht-kontextfreien Komplements.

Wir können es aber aus den Beweisen ableiten:

- Seien $\mathbf{L_1} = \{a^i b^i c^k \mid i \geq 0, k \geq 0\}$ und $\mathbf{L_2} = \{a^i b^k c^k \mid i \geq 0, k \geq 0\}$
- $\mathbf{L_1} \cap \mathbf{L_2} = \overline{\overline{\mathbf{L_1}} \cup \overline{\mathbf{L_2}}} = \{a^i b^i c^i \mid i \geq 0\} = \mathbf{L}$ ist nicht kontextfrei
- $\overline{\mathbf{L_1}} \cup \overline{\mathbf{L_2}}$ ist die Vereinigung der folgenden Typ-2-Sprachen:
  - $\{a, b, c\}^* \circ \{ba, ca, cb\} \circ \{a, b, c\}^*$ (falsche Reihenfolge)
  - $\{a\}^+ \circ \{a^i b^i \mid i \geq 0\} \circ \{c\}^*$ (zu viele a für $\mathbf{L_1}$)
  - $\{a^i b^i \mid i \geq 0\} \circ \{b\}^+ \circ \{c\}^*$ (zu viele b für $\mathbf{L_1}$)
  - $\{a\}^* \circ \{b\}^+ \circ \{b^i c^i \mid i \geq 0\}$ (zu viele b für $\mathbf{L_2}$)
  - $\{a\}^* \circ \{b^i c^i \mid i \geq 0\} \circ \{c\}^+$ (zu viele c für $\mathbf{L_2}$)
- $\overline{\mathbf{L_1}} \cup \overline{\mathbf{L_2}}$ ist daher kontextfrei, aber ihr Komplement nicht

Zusammenfassung und Ausblick

Kontextfreie Sprachen sind abgeschlossen unter Vereinigung, Konkatenation und Kleene-Stern

Kontextfreie Sprachen sind nicht abgeschlossen unter Komplement und Schnitt

Abschlüsse beruhen auf Grammatikoperationen, die man auf beliebige Grammatiken anwenden könnte, aber nur der Abschluss unter $\cup$ für Typ 0 und Typ 1 kann so gezeigt werden.

Offene Fragen:

- Haben kontextfreie Sprachen ein Berechnungsmodell?
- Welche Probleme auf kontextfreien Grammatiken kann man lösen?
- Was gibt es zu Typ 1 und Typ 0 zu sagen?