Prof. Javier Esparza
Philipp Czerner, Kevin Kappelmann, Valentin Krasotin

Technische Universität München
Lehrstuhl für Theoretische Informatik

Einführung in die Theoretische Informatik

Sommersemester 2024 – Übungsblatt 6

- Das Übungsblatt ist in zwei Teile gegliedert: den Vorbereitungsteil, den Sie vor der Übung selbstständig bearbeiten sollen, und den Übungs-/Nachbereitungsteil, der Aufgaben enthält, die in der Übung besprochen werden und von Ihnen anschließend zur Nachbereitung verwendet werden können.

Vorbereitung ($\rightarrow$ vor der Übung selbständig zu bearbeiten)

Vorbereitungsaufgabe Ü6.1. (*Wichtige Begriffe*)

Überprüfen Sie, dass Sie die folgenden Begriffe oder Notationen korrekt definieren können.

- nützlich, erzeugend, erreichbar (Symbole)
- Chomsky-Normalform
- Greibach-Normalform
- Pumping-Lemma für kontextfreie Sprachen

Vorbereitungsaufgabe Ü6.2. (*Automata Tutor: "Chomsky-Normalform"*)

Lösen Sie die Aufgaben Ü6.2 (a–b) auf Automata Tutor.

Hinweis: Automata Tutor überprüft nur, ob Ihre Grammatik in CNF ist und die richtige Sprache erzeugt. Sie selbst sollen die Grammatik jedoch nach Verfahren der Vorlesung in CNF überführen.

Übung und Nachbereitung

Übungsaufgabe Ü6.3. (*Express Yourself (Whoa Do It)*)

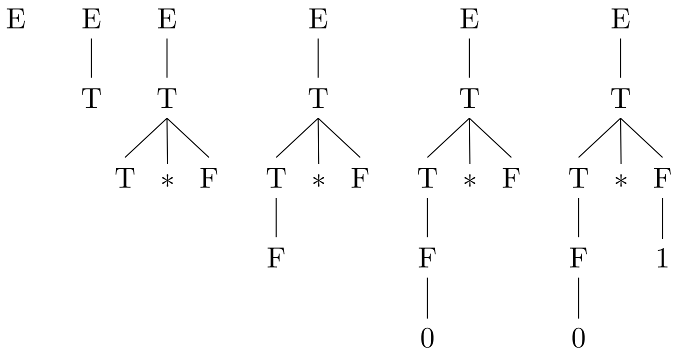
Wie in der Vorlesung besprochen, entspricht jede Ableitung $S \rightarrow \alpha_1 \rightarrow \cdots \rightarrow \alpha_n$ bezüglich einer kontextfreien Grammatik $G = (V, \Sigma, P, S)$ einem Syntaxbaum. Jedes Vorkommnis eines Zeichens $\mathbb{X} \in (V \cup \Sigma \cup \{\varepsilon\})$ in der Ableitung entspricht dabei einem Knoten des Syntaxbaums. Der Baum entsteht dabei induktiv:

- Die Wurzel des Baums ist das Startsymbol S.
- Wird in der Wortableitung auf das Vorkommnis eines Nichtterminals A eine Regel $A \rightarrow \mathbb{X}_1 \dots \mathbb{X}_n$ mit $\mathbb{X}_i \in (V \cup \Sigma \cup \{\varepsilon\})$ angewandt, so werden dem dazugehörigen Knoten entsprechende Kinder $\mathbb{X}_1 \dots \mathbb{X}_n$ hinzugefügt.

Betrachte beispielsweise die Grammatik $G_1 = (\{\mathrm{E}, \mathrm{T}, \mathrm{F}\}, \{(,), 0, 1, +, *\}, P_1, \mathrm{E})$ für arithmetische Ausdrücke (aus der Vorlesung, Beispiel 4.1) mit folgenden Produktionen P_1:

$$\mathrm{E} \rightarrow \mathrm{T} \mid \mathrm{E} + \mathrm{T} \qquad \mathrm{T} \rightarrow \mathrm{F} \mid \mathrm{T} * \mathrm{F} \qquad \mathrm{F} \rightarrow 0 \mid 1 \mid (\mathrm{E})$$

Für die Ableitung $\mathrm{E} \to \mathrm{T} \to \mathrm{T} * \mathrm{F} \to \mathrm{F} * \mathrm{F} \to 0 * \mathrm{F} \to 0 * 1$ entstehen schrittweise die folgenden Syntaxbäume:



Für einen Syntaxbaum eines Wortes w produziert durch G_1 können wir die Semantik (die "mathematische Bedeutung") rekursiv berechnen:

$$\left[\!\!\left[\begin{array}{c}\mathrm{E}\\ \mid \\ \mathcal{B}\end{array}\right]\!\!\right] = [\![\mathcal{B}]\!] \qquad \left[\!\!\left[\begin{array}{c}\mathrm{E}\\ \diagup \mid \diagdown \\ \mathcal{B}_1 \; + \; \mathcal{B}_2\end{array}\right]\!\!\right] = [\![\mathcal{B}_1]\!] + [\![\mathcal{B}_2]\!]$$

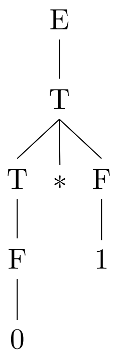
$$\left[\!\!\left[\begin{array}{c}\mathrm{T}\\ \mid \\ \mathcal{B}\end{array}\right]\!\!\right] = [\![\mathcal{B}]\!] \qquad \left[\!\!\left[\begin{array}{c}\mathrm{T}\\ \diagup \mid \diagdown \\ \mathcal{B}_1 \; * \; \mathcal{B}_2\end{array}\right]\!\!\right] = [\![\mathcal{B}_1]\!] * [\![\mathcal{B}_2]\!]$$

$$\left[\!\!\left[\begin{array}{c}\mathrm{F}\\ \mid \\ \mathcal{B}\end{array}\right]\!\!\right] = [\![\mathcal{B}]\!] \qquad \left[\!\!\left[\begin{array}{c}\mathrm{F}\\ \diagup \mid \diagdown \\ (\; \mathcal{B} \;)\end{array}\right]\!\!\right] = [\![\mathcal{B}]\!] \qquad [\![0]\!] = 0 \qquad [\![1]\!] = 1$$

Dabei stehen $\mathcal{B}, \mathcal{B}_1, \mathcal{B}_2$ für beliebige Teilbäume.

Beachten Sie: die Semantik ist hier eine natürliche Zahl, die sie insbesondere mittels arithmetischen Regeln vereinfachen dürfen. Beispielsweise gilt $0 + 1 = 1$ für natürliche Zahlen (semantische Gleichheit), aber $0 + 1 \neq 1$ für Wörter über dem Alphabet Σ (syntaktische Gleichheit).

(a) Bestimmen Sie schrittweise die Semantik des folgenden Syntaxbaums B durch Anwenden der entsprechenden Regeln.



(b) Existiert für das Wort $0 * 1$ ein weiterer Syntaxbaum B', sodass $[\![B]\!] \neq [\![B']\!]$?

(c) Existiert ein Wort $w' \neq 0 * 1$ mit Syntaxbaum B', sodass $[\![B]\!] = [\![B']\!]$?

(d) Betrachte nun die Grammatik $G_2 = (\{S\}, \{\text{if}, \text{then}, \text{else}, 0, 1, +, *, =\}, P_2, S\}$ mit folgenden Produktionen P_2:

$$S \to 0 \mid 1 \mid S + S \mid S * S \mid \text{if } S = S \text{ then } S \text{ else } S$$

Geben Sie, analog zur Grammatik G_1, eine mathematisch sinnvolle Semantik für Syntaxbäume bezüglich G_2 rekursiv an.

(e) Bestimmen Sie einen Syntaxbaum und dessen Semantik für das Wort
$1 + \text{if } 1 = 0 \text{ then } 0 \text{ else } 1$

(f) Existiert ein Wort w mit Syntaxbäumen B, B' bezüglich G_2, sodass $[\![B]\!] \neq [\![B']\!]$?

Übungsaufgabe Ü6.4. (*Pumping-Lemma für Kontextfreie Sprachen*)

Zeigen Sie unter Verwendung des Pumping Lemmas für kontextfreie Sprachen, dass die folgende Sprache nicht kontextfrei ist:

$$L \coloneqq \{a^n b^m c^l \mid n, m, l \in \mathbb{N} \land n > m \land n > l\}$$

Übungsaufgabe Ü6.5. (*Chomsky-Normalform*)

Betrachten Sie die folgende Grammatik G:

$$S \to ASA \mid aB \qquad A \to B \mid S \qquad B \to b \mid \varepsilon$$

(a) Überführen Sie die Grammatik in Chomsky-Normalform.

(b) Erklären Sie in eigenen Worten, wie Sie nach Überführen einer Grammatik in CNF überprüfen können, dass Sie keine Fehler gemacht haben.[1]

Zusätzliche Übungsaufgabe Ü6.6. (*Fitnessstudio*)

Diese Aufgabe können Sie zusätzlich als Nachbereitung lösen. Es wird kein neuer Inhalt in dieser Aufgabe behandelt.

Wir betrachten Pfeil-Sprachen über dem Alphabet $\Sigma = \{\uparrow, \downarrow, \leftarrow, \rightarrow\}$. Wir interpretieren dabei ein Wort $w \in \Sigma^*$ als einen Pfad in einem 2D-Gitter.

(a) Geben Sie eine formale Definition für die folgenden beiden Sprachen an:

(1) Pfade, die in den Ursprung zurückkehren.

(2) Pfade, die „in großer Kurve umkehren", d.h. *beliebig weit* nach rechts fahren, dann *noch weiter* entweder nach oben oder unten gehen und letztlich wieder umkehren und *noch weiter* nach links fahren. Diese Pfade enden dann links vom Ursprung.

(b) Zeigen Sie mit Hilfe des Pumping-Lemma für kontextfreie Sprachen, dass diese Sprachen nicht kontextfrei sind.

Lösungsskizze. Für diese Aufgabe gibt es eine Video-Lösung.

(a) (1)

$$L_a = \{w \in \Sigma^* \mid |w|_\uparrow = |w|_\downarrow \land |w|_\leftarrow = |w|_\rightarrow\}$$

(2)

$$L_b = \bigcup_{i<j<k} L(\rightarrow^i (\uparrow^j \mid \downarrow^j) \leftarrow^k)$$

[1] Hier geht es nicht um einen formalen Beweis, dass die beiden Sprachen gleich sind, sondern um eine Strategie, wie Sie bei Aufgaben wie im vorherigen Aufgabenteil Fehler vermeiden.

(b) (1)

- Wir nehmen an, dass L_a kontextfrei ist und führen diese Annahme zum Widerspruch.
- Sei $n \geq 1$ eine Pumping-Lemma-Zahl für die kontextfreie Sprache L_a.
- Dann gilt $z :=\rightarrow^n\uparrow^n\leftarrow^n\downarrow^n \in L_a$ und $|z| \geq n$.
- Gemäß Pumping-Lemma gibt es dann also eine Zerlegung $z = uvwxy$ mit Wörtern $u, v, w, x, y \in \Sigma^*$ und den folgenden Eigenschaften:

$$(1)\ vx \neq \varepsilon \qquad (2)\ |vwx| \leq n \qquad (3)\ \forall i \in \mathbb{N}.\ uv^iwx^iy \in L_a$$

- Aufgrund von (1) unterscheiden wir die folgenden Fälle:
  - $|vx|_\rightarrow > 0$: Wegen (2) gilt $|vx|_\leftarrow = 0$. Allerdings gilt auch:
  $$|uv^2wx^2y|_\rightarrow = n + |vx|_\rightarrow > n = |uv^2wx^2y|_\leftarrow$$
  Dies ist ein Widerspruch zu (3).
  - $|vx|_\uparrow > 0$: Wegen (2) gilt $|vx|_\downarrow = 0$. Allerdings gilt auch:
  $$|uv^2wx^2y|_\uparrow = n + |vx|_\uparrow > n = |uv^2wx^2y|_\downarrow$$
  Dies ist ein Widerspruch zu (3).
  - $|vx|_\leftarrow > 0$: Wegen (2) gilt $|vx|_\rightarrow = 0$. Allerdings gilt auch:
  $$|uv^2wx^2y|_\rightarrow = n < n + |vx|_\leftarrow = |uv^2wx^2y|_\leftarrow$$
  Dies ist ein Widerspruch zu (3).
  - $|vx|_\downarrow > 0$: Wegen (2) gilt $|vx|_\uparrow = 0$. Allerdings gilt auch:
  $$|uv^2wx^2y|_\uparrow = n < n + |vx|_\downarrow = |uv^2wx^2y|_\downarrow$$
  Dies ist ein Widerspruch zu (3).
- Da jeder Fall zu einem Widerspruch führt und die obigen Fälle alle möglichen Zerlegungen abdecken, kann die ursprüngliche Annahme nicht gelten. Also ist L_a nicht kontextfrei.

(2)

- Wir nehmen an, dass L_b kontextfrei ist und führen diese Annahme zum Widerspruch.
- Sei $n \geq 1$ eine Pumping-Lemma-Zahl für die kontextfreie Sprache L_b.
- Dann gilt $z :=\rightarrow^n\uparrow^{n+1}\leftarrow^{n+2} \in L_b$ und $|z| \geq n$.
- Gemäß Pumping-Lemma gibt es dann also eine Zerlegung $z = uvwxy$ mit Wörtern $u, v, w, x, y \in \Sigma^*$ und den folgenden Eigenschaften:

$$(1)\ vx \neq \varepsilon \qquad (2)\ |vwx| \leq n \qquad (3)\ \forall i \in \mathbb{N}.\ uv^iwx^iy \in L_b$$

- Zuerst informell: Da $|vwx| \leq n$, kann vwx nur von der Form $\rightarrow^*\uparrow^*$ oder $\uparrow^*\leftarrow^*$ sein. Wegen $|vx| > 0$ muss mindestens ein Pfeil gepumpt werden. Gilt $|vx|_\rightarrow > 0$, dann können wir die Anzahl der $\rightarrow$ über die Anzahl der $\leftarrow$ pumpen, enthält vx keinen $\rightarrow$ aber mindestens ein $\uparrow$, so kann man die Anzahl der $\rightarrow$ auf höchstens n reduzieren, indem man vx entfernt. Andernfalls besteht vx nur aus $\leftarrow$, dann aus mindestens einem, so dass man durch Entfernen von vx die Anzahl der $\leftarrow$ auf $n+1$ oder weniger reduziert wird, man also höchstens so viele $\leftarrow$ wie $\uparrow$ hat.

- Formal: Aufgrund von (1) unterscheiden wir die folgenden Fälle:
  - $|vx|_{\rightarrow} > 0$: Wegen (2) gilt $|vx|_{\leftarrow} = 0$. Allerdings gilt auch:

    $$|uv^3wx^3y|_{\rightarrow} = n + 2|vx|_{\rightarrow} \geq n + 2 = |uv^3wx^3y|_{\leftarrow}$$

    Dies ist ein Widerspruch zu (3).
  - $|vx|_{\rightarrow} = 0$ und $|vx|_{\uparrow} > 0$: Dann gilt:

    $$|uv^0wx^0y|_{\uparrow} = n + 1 - |vx|_{\uparrow} \leq n = |uv^0wx^0y|_{\rightarrow}$$

    Daher ist $uv^0wx^0y \notin L_b$, ein Widerspruch zu (3).
  - $|vx|_{\rightarrow} = 0$ und $|vx|_{\uparrow} = 0$: Dann muss $|vx|_{\leftarrow} > 0$ gelten, und es folgt:

    $$|uv^0wx^0y|_{\leftarrow} = n + 2 - |vx|_{\leftarrow} \leq n + 1 = |uv^0wx^0y|_{\uparrow}$$

    Daher ist $uv^0wx^0y \notin L_b$, ein Widerspruch zu (3).

  Da jeder Fall zu einem Widerspruch führt und die obigen Fälle alle möglichen Zerlegungen abdecken, kann die ursprüngliche Annahme nicht gelten. Also ist L_b nicht kontextfrei.