Tutorium
Automaten und Formale Sprachen
Teil IV.

Christopher Blöcker, B. Sc.
inf9900@fh-wedel.de

SS 2012

Definition: Mehrdeutigkeit

Eine Grammatik ist mehrdeutig, wenn es ein Wort gibt, für das unterschiedliche Ableitungsbäume existieren.

Ob eine Grammatik mehrdeutig ist, ist **im Allgemeinen nicht entscheidbar**. Es existiert also kein Algorithmus, der zu jeder Grammatik entscheiden kann, ob sie mehrdeutig ist.

Dennoch ist es unter Vorgabe einer konkreten Grammatik möglich, zu entschieden, ob diese mehrdeutig ist oder nicht.

Nicht-Determinismus in Grammatiken

Grammatiken sind nicht-deterministisch, da die angewendeten Produktionsregeln zufällig ausgewählt werden.

Soll eine Grammatik deterministisch sein, so darf es immer nur eine Produktionsregel zur Auswahl geben.

Die Chomsky-Hierarchie

Noam Chomsky hat den Begriff der Chomsky-Hierarchie geprägt und die folgenden 4 Sprachklassen definiert:

- Typ-3 : Reguläre Sprachen $\mathcal{L}_3$
- Typ-2 : Kontext-Freie Sprachen $\mathcal{L}_2$
- Typ-1 : Kontext-Sensitive Sprachen $\mathcal{L}_1$
- Typ-0 : Rekursiv Aufzählbare Sprachen $\mathcal{L}_0$

Es gilt:

$$\mathcal{L}_3 \subsetneq \mathcal{L}_2 \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0$$

Jede dieser Sprachklassen lässt sich durch Grammatiken des jeweiligen Types erzeugen. Dabei unterliegen die Produktionen gewissen Restriktionen.

Anmerkung

Es gibt (überabzählbar viele) Sprachen, die **nicht einmal** rekursiv aufzählbar sind!

Frage

Wie lässt sich feststellen, ob

$$\mathcal{L} \in \mathcal{L}_2?$$

Antwort

Eine Sprache ist vom Typ-2, wenn sie von einem Kellerautomaten akzeptiert oder von einer kontext-freien Grammatik erzeugt wird. Für den Nachweis $\mathcal{L} \in \mathcal{L}_2$ ist also ein Kellerautomat $\mathcal{A}$ oder eine Typ-2-Grammatik $\mathcal{G}$ anzugeben mit

$$\mathcal{L}_{\mathcal{A}} = \mathcal{L} \text{ bzw. } \mathcal{L}_{\mathcal{G}} = \mathcal{L}.$$

Außerdem lassen sich kontext-freie Grammatiken in nichtdeterministische(!) Kellerautomaten (**NKA**) überführen.

Definition

Ein deterministischer Kellerautomat ist ein 6-Tupel.

$$\mathcal{A} = (\mathcal{Q}, \Sigma, \Gamma, \delta, q_0, \mathcal{F})$$

mit

$\mathcal{Q}$	endliche, nichtleere Menge von Zuständen
Σ	endliche, nichtleere Menge, das Eingabealphabet mit $\Sigma \cap \mathcal{Q} = \emptyset$ und $\Sigma_\epsilon = \Sigma \cup \{\epsilon\}$
Γ	endliche, nichtleere Menge, das Kelleralphabet mit $\perp \in \Gamma$ und $\Gamma_\epsilon = \Gamma \cup \{\epsilon\}$
δ	Überführungsfunktion mit $\delta : \mathcal{Q} \times \Sigma_\epsilon \times \Gamma_\epsilon \to \mathcal{Q} \times \Gamma_\epsilon^*$
q_0	Startzustand
$\mathcal{F}$	Menge von Endzuständen mit $\mathcal{F} \subseteq \mathcal{Q}$

Definition

Ein nichtdeterministischer Kellerautomat ist ein 6-Tupel.

$$\mathcal{A} = (\mathcal{Q}, \Sigma, \Gamma, \delta, q_0, \mathcal{F})$$

mit

$\mathcal{Q}$	endliche, nichtleere Menge von Zuständen
Σ	endliche, nichtleere Menge, das Eingabealphabet mit $\Sigma \cap \mathcal{Q} = \emptyset$ und $\Sigma_\epsilon = \Sigma \cup \{\epsilon\}$
Γ	endliche, nichtleere Menge, das Kelleralphabet mit $\perp \in \Gamma$ und $\Gamma_\epsilon = \Gamma \cup \{\epsilon\}$
δ	Überführungsrelation mit $\delta \subseteq \mathcal{Q} \times \Sigma_\epsilon \times \Gamma_\epsilon \times \mathcal{Q} \times \Gamma_\epsilon^*$
q_0	Startzustand
$\mathcal{F}$	Menge von Endzuständen mit $\mathcal{F} \subseteq \mathcal{Q}$

Typ-2 Grammatik

Sei

$$\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{R}, \mathcal{S}).$$

$\mathcal{G}$ ist vom Typ-2, wenn die Produktionen folgende Form haben:

$$A \quad \rightarrow \quad w$$

mit

- $A \in \mathcal{N}$
- $w \in (\mathcal{N} \cup \mathcal{T})^*$

Frage

Wie lässt sich feststellen, ob

$$\mathcal{L} \notin \mathcal{L}_2?$$

Antwort

Mit Hilfe des Pumping-Lemmas für kontext-freie Sprachen.

Wenn die Bedingungen des Pumping-Lemmas nicht erfüllt sind, so kann die betroffene Sprache nicht kontext-frei sein.

$$\overline{\mathrm{PL}_{\mathrm{kfS}}(\mathcal{L})} \to \mathcal{L} \notin \mathcal{L}_2.$$

Das Pumping-Lemma für kontext-freie Sprachen

Sei $\mathcal{L} \subseteq \Sigma^*$ eine kontext-freie Sprache. Dann gibt es eine natürliche Zahl $p \in \mathbb{N}$ derart, dass sich jedes Wort $w \in \mathcal{L}$ mit $|w| \geq p$ zerlegen lässt in fünf Teilworte

$$w = uvxyz$$

mit

1. $|vy| > 0$
2. $|vxy| \leq p$
3. $\bigwedge_{i \in \mathbb{N}} uv^i xy^i z \in \mathcal{L}$

Vorsicht!

Es gilt

$$\mathcal{L} \in \mathcal{L}_2 \to \mathrm{PL}_{\mathrm{kfS}}(\mathcal{L}).$$

Die Umkehrung gilt jedoch **nicht**! Die folgende Aussage ist also **falsch**:

$$\lightning\ \mathrm{PL}_{\mathrm{kfS}}(\mathcal{L}) \to \mathcal{L} \in \mathcal{L}_2.\ \lightning$$

Aufgaben

Prüfen Sie auf Kontext-freiheit:

1. $L_1 = \left\{a^i b^j \mid i, j \in \mathbb{N} \land i \neq j\right\}$
2. $L_2 = \left\{a^i b^j \mid i, j \in \mathbb{N} \land i = j^2\right\}$

Herangehensweise

- Angeben eines Kellerautomaten oder einer kontext-freien Grammatik für den Nachweis $L \in \mathcal{L}_2$
- Benutzung des Pumping-Lemma für kontext-freie Sprachen für den Nachweis $L \notin \mathcal{L}_2$

$L_1 = \{a^i b^j \mid i,j \in \mathbb{N} \wedge i \neq j\}$

Wir müssen eine Grammatik finden, bei der stets entweder mehr a's ober mehr b's in den erzeugten Wörtern auftreten. Außerdem darf kein b vor einem a stehen.

Sei $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{R}, \mathcal{S})$ mit

$$\begin{array}{lcl} \mathcal{N} & = & \{S, A, B, C\} \\ \mathcal{T} & = & \{a, b\} \\ \mathcal{R} & : & S \rightarrow A \mid B \\ & & A \rightarrow aA \mid aC \\ & & B \rightarrow Bb \mid Cb \\ & & C \rightarrow aCb \mid \epsilon \end{array}$$

$\mathcal{G}$ erzeugt $L_1 \rightarrow L_1 \in \mathcal{L}_2$ $\square$

Transformation $\mathcal{G}_{kfS} \rightarrow$ NKA

Sei $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{R}, \mathcal{S})$.
Dann lässt sich daraus ein **NKA** $\mathcal{A}$ mit $\mathcal{L}_\mathcal{A} = \mathcal{L}_\mathcal{G}$ bestimmen mit

$$\mathcal{A} = (\mathcal{Q}, \Sigma, \Gamma, \delta, q_0, \mathcal{F})$$

wobei

$$\begin{array}{lcl} \mathcal{Q} & \leftarrow & \{q_0, q_1, q_2\} \\ \Sigma & \leftarrow & \mathcal{T} \\ \Gamma & \leftarrow & \mathcal{T} \cup \mathcal{N} \cup \{\perp\} \\ \delta & \leftarrow & \{(q_0, \epsilon, \epsilon, q_1, S\perp), (q_1, \epsilon, \perp, q_2, \epsilon)\} \\ & \cup & \{(q_1, \epsilon, A, q_1, w) \mid A \rightarrow w \in \mathcal{R}\} \\ & \cup & \{(q_1, a, a, q_1, \epsilon) \mid a \in \mathcal{T}\} \\ \mathcal{F} & \leftarrow & \{q_2\} \end{array}$$

Alternativer, **fehlerhafter!** Beweis:

$L_1 = \{a^i b^j \mid i, j \in \mathbb{N} \wedge i \neq j\}$

Da die Menge der kontext-freien Sprachen abgeschlossen ist bezüglich Komplementbildung, kann $L_1 \in \mathcal{L}_2$ gezeigt werden, wenn bewiesen werden kann, dass $\overline{L_1} \in \mathcal{L}_2$.

$$\overline{L_1} = \{a^i b^j \mid i, j \in \mathbb{N} \wedge i = j\} = \{a^n b^n \mid n \in \mathbb{N}\}$$

Sei $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{R}, \mathcal{S})$ mit

$$\begin{array}{lcl} \mathcal{N} & = & \{S\} \\ \mathcal{T} & = & \{a, b\} \\ \mathcal{R} & : & S \to aSb \mid \epsilon \end{array}$$

$\mathcal{G}$ erzeugt $\overline{L_1} \to \overline{L_1} \in \mathcal{L}_2 \to L_1 \in \mathcal{L}_2$ □

Fehleraufdeckung:

$L_1 = \{a^i b^j \mid i,j \in \mathbb{N} \wedge i \neq j\}$

- $\surd$ Die Menge der kontext-freien Sprachen ist abgeschlossen bezüglich Komplementbildung
- $\surd$ $L_1 \in \mathcal{L}_2 \Leftrightarrow \overline{L_1} \in \mathcal{L}_2$
- $\lightning$ $\overline{L_1} = \{a^i b^j \mid i,j \in \mathbb{N} \wedge i = j\} = \{a^n b^n \mid n \in \mathbb{N}\}$

Komplementärsprache

Das Komplement einer Sprache, $\overline{L}$ besteht aus den Wörtern, die **nicht** in L sind.

$$\overline{L} = \{w \in \Sigma^* \mid w \notin L\} = \Sigma^* \setminus L.$$

$$\Rightarrow L' = \{a^i b^i \mid i \in \mathbb{N}\} \text{ ist nur eine Teilmenge von } \overline{L_1}.$$

$$\overline{L_1} = \{w \in \{a,b\}^* \mid w \text{ hat nicht die Form } a^i b^j \text{ mit } i \neq j\}$$

$L_2 = \left\{ a^i b^j \mid i, j \in \mathbb{N} \wedge i = j^2 \right\}$

Die Wörter von L_2 haben die Form $w = a^{j^2} b^j$.

Wenn $L_2 \in \mathcal{L}_2$, dann muss das $\mathrm{PL}_{\mathrm{kfS}}$ gelten.

$$\text{Sei } w = a^{p^2} b^p.$$

Wir müssen nun das Wort zerteilen in $w = uvxyz$, wobei beachtet werden muss, dass nur Wörter zu L gehören, bei denen kein a vor einem b steht. Außerdem muss bei zunehmender Anzahl von a's auch die Anzahl der b's zunehmen und umgekehrt.

Folglich kann v nur aus a's und y nur aus b's bestehen.

$$\text{Sei die Zerteilung } w = a^{p^2} b^p = \underbrace{a^{p^2-m-n}}_{u} \cdot \underbrace{a^m}_{v} \cdot \underbrace{a^n b^r}_{x} \cdot \underbrace{b^s}_{y} \cdot \underbrace{b^{p-r-s}}_{z}.$$

$L_2 = \left\{a^i b^j \mid i,j \in \mathbb{N} \wedge i = j^2\right\}$

Beim Pumpem erhalten wir die Wörter

$$w_i = a^{p^2-m-n} \cdot a^{i \cdot m} \cdot a^n b^r \cdot b^{i \cdot s} \cdot b^{p-r-s}.$$

Damit $w_i \in L_2$, muss gelten

$$\begin{aligned}
p^2 - m - n + i \cdot m + n &= (p - r - s + i \cdot s + r)^2 \\
p^2 - m + i \cdot m &= (p - s + i \cdot s)^2 \\
p^2 + m\,(i-1) &= (p + s\,(i-1))^2 \\
p^2 + (i-1)\,m &= p^2 + 2ps\,(i-1) + s^2\,(i-1)^2 \\
(i-1)\,m &= 2ps\,(i-1) + s^2\,(i-1)^2 \\
m &= 2ps + s^2\,(i-1) \\
i &= \frac{m - 2ps}{s^2} + 1
\end{aligned}$$

Damit $w_i \in L_2$, muss i also konstant sein.

$\Rightarrow w$ lässt sich nicht für jedes i pumpen $\rightarrow L_2 \notin \mathcal{L}_2$ $\square$