3 Kontextfreie Sprachen

Idee/Ziel: Beschreibung typischer Programmiersprachen und
des Verhaltens von rekursiven Programmen

Beispiel 3.1.

Mithilfe der folgenden Regeln werden Ausdrücke mit $+$, $-$, $\cdot$, $/$ und Konstanten 0 und 1 beschrieben.

$$
\begin{array}{lll}
\langle \mathsf{expr} \rangle \to \langle \mathsf{expr} \rangle + \langle \mathsf{expr} \rangle & \langle \mathsf{term} \rangle \to \langle \mathsf{term} \rangle \cdot \langle \mathsf{term} \rangle & \langle \mathsf{factor} \rangle \to 0 \\
\langle \mathsf{expr} \rangle \to \langle \mathsf{expr} \rangle - \langle \mathsf{expr} \rangle & \langle \mathsf{term} \rangle \to \langle \mathsf{term} \rangle / \langle \mathsf{term} \rangle & \langle \mathsf{factor} \rangle \to 1 \\
\langle \mathsf{expr} \rangle \to \langle \mathsf{term} \rangle & \langle \mathsf{term} \rangle \to \langle \mathsf{factor} \rangle & \langle \mathsf{factor} \rangle \to (\langle \mathsf{expr} \rangle)
\end{array}
$$

Dies soll jetzt verallgemeinert werden.

3.1 Kontextfreie Grammatiken

Definition. Eine *kontextfreie Grammatik* ist ein Tupel $G = (N, \Sigma, S, P)$ mit folgenden Komponenten:

- N ist eine endliche Menge von *Nichtterminalen*,
- Σ ist ein Alphabet von *Terminalen* mit $N \cap \Sigma = \emptyset$,
- $S \in N$ ist das *Startsymbol* und
- $P \subseteq N \times (N \cup \Sigma)^*$ ist eine endliche Menge von *Regeln*.

Eine *rechtslineare Grammatik* ist eine kontextfreie Grammatik (N, Σ, S, P) mit $P \subseteq N \times \Sigma N \cup \{\varepsilon\}$.

Beispiel 3.1 (Fortsetzung)
$N = \{\langle\mathsf{expr}\rangle, \langle\mathsf{term}\rangle, \langle\mathsf{factor}\rangle\}$
$\Sigma = \{+, -, \cdot, /, 0, 1, (,)\}$
$S = \langle\mathsf{expr}\rangle$
P ist die Menge der angegebenen Regeln (wobei $A \to v$ für (A, v) steht).
Dies ist keine rechtslineare Grammatik (z.B. wegen der Regel $\langle\mathsf{term}\rangle \to \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle$).

Konvention Elemente von N: $\langle\mathsf{expr}\rangle$, A, B, ...
Elemente von Σ: a, b, ...

Definition. Sei $G = (N, \Sigma, S, P)$ eine kontextfreie Grammatik und $u, v \in (N \cup \Sigma)^*$.

- Das Wort v ist *in einem Schritt* aus dem Wort u *ableitbar*, wenn es Regel $A \to y$ in P und Wörter $x, z \in (N \cup \Sigma)^*$ gibt mit $u = xAz$ und $v = xyz$. Hierfür schreiben wir $u \Rightarrow_G v$.

- Das Wort v ist *(in beliebig vielen Schritten)* aus dem Wort u *ableitbar*, wenn es $n \in \mathbb{N}$ und Wörter $u_0, u_1, \ldots, u_n$ gibt mit $u = u_0 \Rightarrow_G u_1 \Rightarrow_G u_2 \cdots \Rightarrow_G u_n = v$. Hierfür schreiben wir $u \Rightarrow_G^* v$ (insbes. gilt mit $n = 0$ auch $u \Rightarrow_G^* u$).

- Die von G *erzeugte Sprache* ist

$$L(G) = \{u \in \Sigma^* \mid S \Rightarrow_G^* u\}\,.$$

- Eine Sprache L heißt *kontextfrei*, wenn es eine kontextfreie Grammatik G gibt mit $L = L(G)$.

- Eine Sprache L heißt *rechtslinear*, wenn es eine rechtslineare Grammatik G gibt mit $L = L(G)$.

Beispiel 3.1 (Fortsetzung)

Wir betrachten wieder die Grammatik mit den folgenden Regeln:

$$
\begin{aligned}
\langle\mathsf{expr}\rangle &\to \langle\mathsf{expr}\rangle + \langle\mathsf{expr}\rangle\\
\langle\mathsf{expr}\rangle &\to \langle\mathsf{expr}\rangle - \langle\mathsf{expr}\rangle\\
\langle\mathsf{expr}\rangle &\to \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{term}\rangle / \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{factor}\rangle\\
\langle\mathsf{factor}\rangle &\to 0\\
\langle\mathsf{factor}\rangle &\to 1\\
\langle\mathsf{factor}\rangle &\to (\langle\mathsf{expr}\rangle)
\end{aligned}
$$

Dann gibt es z.B. die folgende Ableitung:

$$
\begin{aligned}
\langle\mathsf{expr}\rangle &\Rightarrow_G \langle\mathsf{term}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/\langle\mathsf{term}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/\langle\mathsf{factor}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle - \langle\mathsf{expr}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{term}\rangle - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle \cdot 0/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G 1 \cdot 0/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G 1 \cdot 0/(1 - 0)
\end{aligned}
$$

Beispiel. Sei $G = (N, \Sigma, S, P)$ die kontextfreie Grammatik mit $N = \{S\}$, $\Sigma = \{a, b\}$ und $P = \{S \to \varepsilon, S \to aSb\}$. Wegen $S \Rightarrow_G \varepsilon$ gilt $\varepsilon \in L(G)$. Außerdem $S \Rightarrow_G aSb \Rightarrow_G aaSbb \cdots \Rightarrow_G a^n S b^n \Rightarrow_G a^n \varepsilon b^n$ für alle $n \geq 1$. Da es keine weiteren Ableitungen gibt, gilt sogar

$$L(G) = \{a^n b^n \mid n \in \mathbb{N}\},$$

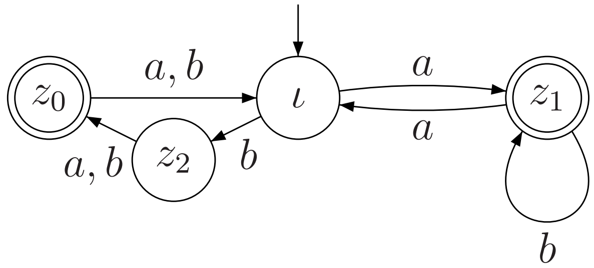
diese Sprache ist also kontextfrei. Im Kapitel 2 haben wir gesehen, daß diese Sprache nicht regulär ist. Es gibt also kontextfreie Sprachen, die nicht regulär sind.

Beispiel. Sei $G = (N, \Sigma, S, P)$ die kontextfreie Grammatik mit $N = \{S\}$, $\Sigma = \{a, b\}$ und $P = \{S \to \varepsilon, S \to aSa, S \to bSb\}$. Dann gilt $L(G) = \{ww^R \mid w \in \Sigma^*\}$. Diese Sprache ist also kontextfrei.

Satz 3.2. *Eine Sprache $L \subseteq \Sigma^*$ ist genau dann regulär, wenn sie rechtslinear ist.*

Beweis durch Beispiel:

Sei zunächst L regulär. Dann existiert ein DFA M mit $L(M) = L$, z.B. sei M der rechts stehende DFA:



Konstruiere eine rechtslineare Grammatik $G = (N, \Sigma, S, P)$ mit $L(G) = L(M) = L$:

- $N = \{z_0, \iota, z_1, z_2\}$,
- $\Sigma = \{a, b\}$,
- $S = \iota$ und

P umfaßt die folgenden Regeln:

$$\begin{array}{lll} \iota \to az_1 & \iota \to bz_2 & \\ z_0 \to a\,\iota & z_0 \to b\,\iota & z_0 \to \varepsilon \\ z_1 \to a\,\iota & z_1 \to bz_1 & z_1 \to \varepsilon \\ z_2 \to az_0 & z_2 \to bz_0 & \end{array}$$

Damit ist die Implikation „$\Longrightarrow$“ am Beispiel gezeigt.

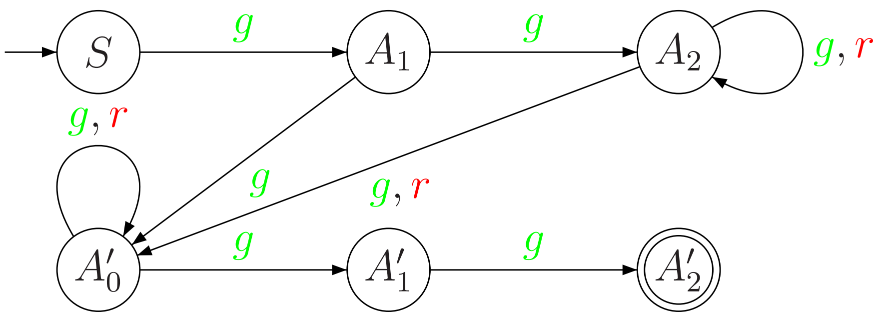
Sei umgekehrt $G = (N, \Sigma, S, P)$ eine rechtslineare Grammatik mit $L(G) = L$, z.B.

- $N = \{S, A_1, A_2, A'_0, A'_1, A'_2\}$,
- $\Sigma = \{g, r\}$ und

P umfaßt die folgenden Regeln:

$$
\begin{array}{rlrl}
S & \to gA_1 & A'_0 & \to gA'_0 \mid rA'_0 \mid gA'_1 \\
A_1 & \to gA_2 \mid gA'_0 & A'_1 & \to gA'_2 \\
A_2 & \to gA_2 \mid rA_2 \mid gA'_0 \mid rA'_0 & A'_2 & \to \varepsilon
\end{array}
$$

Hieraus konstruieren wir den folgenden NFA M mit $L(M) = L(G) = L$:



Also ist L nach Satz 2.10 regulär $\square$

Bemerkung. Wir erhalten:
L regulär
$\Longrightarrow L$ ist Sprache einer rechtslinearen Grammatik
$\Longrightarrow L$ ist Sprache einer kontextfreien Grammatik
$\Longrightarrow L$ ist kontextfrei,
d.h. jede reguläre Sprache ist kontextfrei.

Aber nicht jede kontextfreie Sprache ist regulär (Bsp.: $\{a^n b^n \mid n \in \mathbb{N}\}$).