Theoretische Informatik 1
Große Übung 5

René Maseli
Prof. Dr. Roland Meyer
TU Braunschweig
Wintersemester 2025/26

Chomsky-Normalform

Es sei die folgende kontextfreie Grammatik G gegeben über $\Sigma = \{c, d\}$.

$$S \to A \mid ABB \qquad A \to Scc \mid \varepsilon \qquad B \to dB \mid c$$

Zur Erinnerung: Eine Grammatik ist in Chomsky-Normalform (CNF), falls alle Produktionen die Form $X \to YZ$ oder $X \to s$ haben. Die Greibach-Normalform (GNF) benötigt alle Produktionen in der Form $X \to sX_1 \ldots X_n$ (mit $X, Y, Z, X_1, \ldots X_n \in N$ und $s \in \Sigma$).

Für die CNF eliminiert man zuerst die ε-Produktionen (sowohl A als auch S können zu ε übersetzt werden). Die resultierende Grammatik G'' erfüllt $\mathcal{L}(G'') = \mathcal{L}(G') \setminus \{\varepsilon\} = \mathcal{L}(G) \setminus \{\varepsilon\}$.

G' : $S \to A \mid ABB \mid \varepsilon \mid BB$ $A \to Scc \mid \varepsilon \mid cc$ $B \to dB \mid c$

G'' : $S \to A \mid ABB \mid BB$ $A \to Scc \mid cc$ $B \to dB \mid c$

Nur $S \to BB$ und $B \to c$ haben schon CNF, der Rest muss angepasst werden, notfalls mit neuen Nichtterminalen. Dadurch erhalten wir eine Grammatik G''' in CNF mit $\mathcal{L}(G''') = \mathcal{L}(G'')$.

G''' : $S \to SE \mid CC \mid AF \mid BB$ $A \to SE \mid CC$ $B \to DB \mid c$ $C \to c$ $D \to d$ $E \to CC$ $F \to BB$

Wortproblem lösen mit dem CYK-Algorithmus

Gegeben die kontextfreie Grammatik G, nutze den CYK-Algorithmus, um zu prüfen, ob *ccdecde* von G erzeugt werden kann.

$$S \to ABC \mid CAB \qquad A \to dC \mid eBC \qquad B \to ED \mid BB \mid EA \qquad C \to c\,.$$

Für den Algorithmus aus der Vorlesung benötigen wir eine Grammatik in Chomsky-Normalform.

$$S \to AF \mid CH \quad A \to DC \mid EF \quad B \to ED \mid BB \mid EA \quad C \to c \quad D \to d \quad E \to e \quad F \to BC \quad H \to AB\,.$$

Die Tabelle wird diagonal-weise gefüllt.

	1	2	3	4	5	6	7
1	E	–	–	A	–	H	AHS
2		E	B	BF	–	B	BF
3			D	A	–	H	HS
4				C	–	–	–
5					E	B	BF
6						D	A
7							C

In der oberen rechten Zelle soll sich genau dann S befinden, wenn das Wort in der Sprache liegt. Hier können wir also $eedcedc \in \mathcal{L}(G') = \mathcal{L}(G)$ schließen.

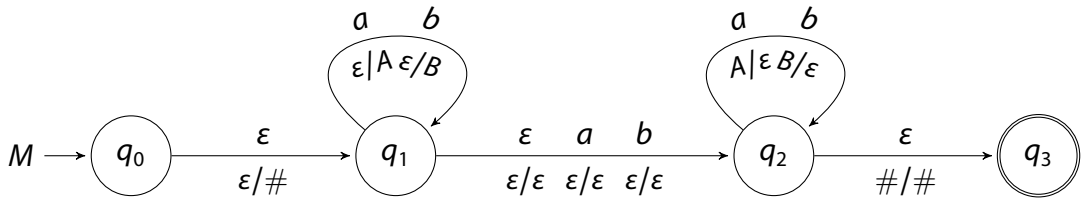
Darüber hinaus lesen wir anhand der Anzahl von S in der ganzen Tabelle, dass 2 nicht-leere Infixe des Wortes in der Sprache liegen: Neben *eedcedc* selbst gibt es noch $dcedc \in \mathcal{L}(G)$. Die oberen Zeile verrät Eigenschaften über Präfixe und die rechteste Spalte steht für die Suffixe des Wortes.

Konstruktion von Pushdown-Automaten

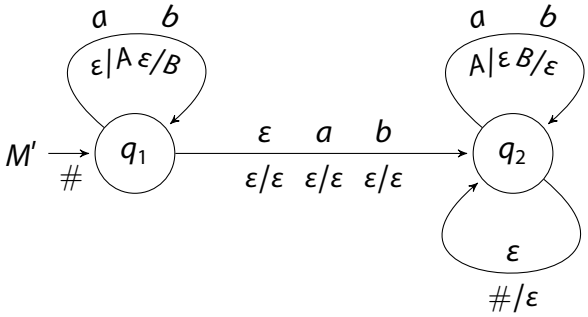
Gesucht ist ein PDA für die Palindrom-Sprache $\{ w \in \{a, b\}^* \mid w = \text{rev}(w) \}$.

Idee: Der erste Hälfte des Wortes wird in den Stack gepusht und beim Verarbeiten der zweiten Hälfte werden die Symbole abgeglichen und gepopt. Dazu nutzt man Stack-Symbole A und B. Der Automat hat keine Möglichkeit, deterministisch die Mitte des Wortes zu erkennen, also muss er sie während des Runs nicht-deterministisch erraten. Wörter können gerade oder ungerade sein, also darf in der Mitte ein Buchstabe konsumiert werden.

Runs, die falsch raten, dürfen aber nicht akzeptieren. Das könnte sonst bei jedem Wort passieren, wenn die geratene Mitte sich als das letzte Symbol der Eingabe herausstellt. Der Stack muss komplett abgebaut werden, und damit der Automat das feststellen kann, braucht man ein weiteres Stack-Symbol $\#$, das nur einmal und nur am Bottom of Stack vorkommt.



Eine Variante, die mit leerem Stack akzeptiert, statt mit akzeptierenden Zuständen wie oben, kann wiefolgt aussehen:



Addendum: Jeder PDA kann in eine kontextfreie Grammatik überführt werden. Die starken Linksableitungen werden dabei Läufe des Automaten simulieren.

Mit einem Nichtterminal fassen wir einen Teil der PDA-Berechnung zusammen: Nichtterminale haben die Form $\langle p, A, q\rangle \in Q \times \Gamma \times Q$ und beschreiben alle Teil-Läufe, die in p mit A auf dem Stack starten, und irgendwann im Zustand q erstmals das darunterliegende Stack-Symbol freilegen, bzw. den Stack leeren.

In der naiven Grammatik sind üblicherweise die meisten Produktionen nutzlos, also entweder nicht erreichbar oder nicht produktiv. Der folgende Algorithmus berechnet einen nennenswert kleineren erreichbaren Teil:

Require: PDA $M = \langle Q, \Sigma, \Gamma, q_0, \#, \delta\rangle$ (akz. mit leerem Stack)
Ensure: $\mathcal{L}(G) = \mathcal{L}(M)$

$P \leftarrow \varnothing$
$T \leftarrow \{ \langle p, A, q\rangle \mid p \xrightarrow[A/\beta]{\sigma} q' \text{ und } p' \xrightarrow[B/\varepsilon]{s} q \}$ (mit den richtigen Beobachtungen kleiner)
$N \leftarrow \{S\}$
$N_{\text{done}} \leftarrow \{S\}$
for $\langle p, A, q\rangle \in T$ **do**
  $P \leftarrow P \cup \{S \to \langle q_0, \#, q\rangle\}$
  $N \leftarrow N \cup \{\langle q_0, \#, q\rangle\}$
end for
while $N \neq N_{\text{done}}$ **do**
  Sei $\langle p, A, q\rangle \in N \setminus N_{\text{done}}$
  for $p \xrightarrow[A/B_n \ldots B_1]{s} q'$ und $\langle p_1, B_1, q_1\rangle \ldots \langle p_n, B_n, q_n\rangle \in T^n$ **do**
    if $p_1 = q'$ und $q_n = q$ und $q_i = p_{i+1}$ für alle $i \in \{1, \ldots, n-1\}$ **then**
      $P \leftarrow P \cup \{\langle p, A, q\rangle \to s \langle p_1, B_1, q_1\rangle \ldots \langle p_n, B_n, q_n\rangle\}$
      $N \leftarrow N \cup \{\langle p_1, B_1, q_1\rangle, \ldots, \langle p_n, B_n, q_n\rangle\}$
    end if
  end for
  $N_{\text{done}} \leftarrow N_{\text{done}} \cup \{\langle p, A, q\rangle\}$
end while
return $G := \langle N, \Sigma, S, P\rangle$

Bemerkung

Achtet darauf, dass die Reihenfolge nach LIFO-Prinzip im Stack verkehrt herum gelesen wird: In $p \xrightarrow[A/B_n \ldots B_1]{s} q'$ wird B_n als erstes gepusht und daher als letztes verarbeitet. B_1 wird als letztes gepusht und muss deshalb zuerst verarbeitet werden.