Sei $M = (Z, \Sigma, \delta, z_0, E)$ ein DFA. (Angenommen sei, dass M keine unerreichbaren Zustände enthält; diese können offensichtlich gestrichen werden.) Die Relation $\approx_M \subseteq Z \times Z$ ist definiert durch:

$$z \approx_M z' \iff \text{für alle } x \in \Sigma^* \text{ gilt } \hat{\delta}(z,x) \in E \Leftrightarrow \hat{\delta}(z',x) \in E.$$

$\approx_M$ ist offensichtlich eine Äquivalenzrelation.

Satz 1.4.5 Sei $M = (Z, \Sigma, \delta, z_0, E)$ ein DFA, $L = \mathcal{L}(M)$. Es ist $|Z| = \mathcal{I}(\sim_L)$ (d.h. M ist minimal) genau dann, wenn für alle $z, z' \in Z$ gilt:

$$z \approx_M z' \iff z = z'.$$

Es gilt:

$$\begin{aligned} z \approx_M z' &\iff \hat{\delta}(z,x) \in E \Leftrightarrow \hat{\delta}(z',x) \in E \text{ für alle } x \in \Sigma^* \\ &\iff \text{für alle } k \in \mathbb{N}_0 : \underbrace{\hat{\delta}(z,x) \in E \Leftrightarrow \hat{\delta}(z',x) \in E \text{ für alle } x \in \Sigma^*, |x| \leq k}_{z\, R_k\, z'}. \end{aligned}$$

wobei für die Relationen R_k, $k \in \mathbb{N}_0$ gilt:

$$\begin{aligned} z\, R_0\, z' &\iff z \in E \Leftrightarrow z' \in E, \\ z\, R_{k+1}\, z' &\iff z\, R_k\, z' \text{ und } \hat{\delta}(z, aw) \in E \Leftrightarrow \hat{\delta}(z', aw) \text{ für alle } a \in \Sigma,\ |w| \leq k \\ &\iff z\, R_k\, z' \text{ und } \delta(z,a)\, R_k\, \delta(z',a) \text{ für alle } a \in \Sigma. \end{aligned}$$

Minimalautomaten-Konstruktion

Sei DFA M (ohne unerreichbare Zustände) gegeben. Der folgende Algorithmus konstruiert einen Minimalautomaten zu M. Als Datenstruktur wird eine Tabelle aller Zustandspaare (z, z') von M mit $z \neq z'$ verwendet.

1. Markiere alle Paare (z, z') mit $z \in E$, $z' \notin E$ oder umgekehrt
   ($*$ Diese Paare sind nicht in R_0 $*$).

2. Wiederhole so lange, bis sich die Tabelle nicht mehr ändert:
   Für jedes unmarkierte Paar (z, z'): Falls es $a \in \Sigma$ gibt, so dass $(\delta(z,a), \delta(z',a))$ (in der Liste vorhanden und) markiert ist, so markiere (z, z')
   ($*$ Sukzessive Markierung der Paare, die nicht in $R_1, R_2, \ldots$ sind; am Ende der Iteration gilt $z \approx_M z'$ für alle unmarkierten Paare (z, z') $*$).

3. Verschmelze alle Zustände z und z', für die (z, z') unmarkiert ist.

1.5 Kontextfreie Sprachen

(A) Syntaxbäume

Eine Ableitung in einer kontextfreien (und damit auch in einer regulären) Grammatik ist graphisch darstellbar durch einen ***Syntaxbaum*** (***Ableitungsbaum***, ***parsing tree***).

Sei $S = x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \ldots \Rightarrow x_n = x$ eine Ableitung von x.

Schrittweiser Aufbau eines Syntaxbaumes SB ("für x")

- SB := ein einziger Knoten S (Wurzel von SB);
- Für $i = 1, \ldots, n$: Wird im Ableitungsschritt $x_{i-1} \Rightarrow x_i$ eine Regel $A \to v_1 \ldots v_m$ ($v_j \in V \cup \Sigma$) angewendet, so füge an den Knoten A von SB die Nachfolgeknoten $v_1 \ldots v_m$ an (im Falle einer Regel $A \to \varepsilon$ einen Nachfolgeknoten ε).

Definition. Eine kontextfreie Grammatik G heißt ***mehrdeutig***, wenn es $x \in \mathcal{L}(G)$ gibt derart, dass für x verschiedene Syntaxbäume existieren. Eine kontextfreie Sprache L heißt ***inhärent mehrdeutig***, wenn jede Grammatik G mit $\mathcal{L}(G) = L$ mehrdeutig ist.

(B) Normalformen

Lemma 1.5.1 Zu jeder ε-produktionsfreien kontextfreien Grammatik gibt es eine äquivalente ε-produktionsfreie kontextfreie Grammatik, die keine Regeln der Form $A \to B$ (mit $A, B \in V$) enthält.

Definition. Eine kontextfreie Grammatik $G = (V, \Sigma, P, S)$ ist in ***Chomsky–Normalform***, wenn sie ε-produktionsfrei ist und wenn alle Regeln von G (mit der eventuellen Ausnahme von $S \to \varepsilon$) von der Form $A \to BC$ oder $A \to a$ sind ($A, B, C \in V$, $a \in \Sigma$).

Satz 1.5.2 Zu jeder kontextfreien Grammatik gibt es eine äquivalente Grammatik in Chomsky-Normalform.

Bemerkungen

1. Die Chomsky-Normalform ist hauptsächlich für theoretische Betrachtungen vorteilhaft.

2. Für "günstiges Parsen" ist eine andere Normalform wichtig:
   Greibach-Normalform: ε-produktionsfrei und alle Regeln (außer $S \to \varepsilon$) sind von der Form $A \to aB_1 \ldots B_k$ ($A, B_i \in V$, $a \in \Sigma$, $k \geq 0$).
   Es gilt: Zu jeder kontextfreien Grammatik gibt es eine äquivalente Grammatik in Greibach-Normalform (sogar mit $0 \leq k \leq 2$).