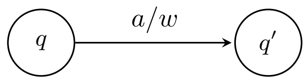
Konečné překladače

Konečný automat můžeme změnit na *konečný překladač*, který nepřijímá jazyk, ale „překládá“ vstupní slovo na výstupní, tedy definuje zobrazení $\Sigma^* \to \Delta^*$. Překladač má hrany



kde $a \in \Sigma$ a $w \in \Delta^*$, které znamenají instrukci „přečti a ze vstupu, změň stav na q' a zapiš w do výstupu“. Konečný překladač je tedy šestice,

$$\mathcal{T} = (Q, \Sigma, q_0, \delta, \Delta, \lambda)$$

kde

- Q je množina stavů,
- Σ je vstupní abeceda,
- q_0 je počáteční stav,
- $\delta : Q \times \Sigma \to Q$ je přechodová funkce;
- Δ je výstupní abeceda a
- $\lambda : Q \times \Sigma \to \Delta^*$ je výstupní funkce.

Slovo w na obrázku výše je $\lambda(q, a)$. Výstupem překladače $\mathcal{T}$ na vstupu $u \in \Sigma^+$ je slovo $\mathcal{T}(u) \in \Delta^*$ definované rekursivně předpisem

$$\mathcal{T}(c) = \lambda(q_0, c),$$
$$\mathcal{T}(uc) = \mathcal{T}(u)\lambda(q_0 \cdot u, c),$$

kde $c \in \Sigma$. Pro jazyk L definujeme

$$\mathcal{T}(L) = \{\mathcal{T}(u) \mid u \in L\} \quad \text{a} \quad \mathcal{T}^{-1}(L) = \{u \mid \mathcal{T}(u) \in L\}.$$

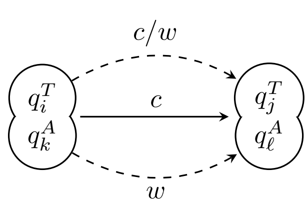
Ukážeme, že regulární jazyky jsou uzavřené na konečné překlady i na jejich vzory.

Věta. Je-li $L \subseteq \Delta^*$ regulární jazyk, je i $\mathcal{T}^{-1}(L) \subseteq \Sigma^*$ regulární.

Důkaz. Nechť $\mathcal{A} = (Q_A, \Delta, q_0^A, \delta_A, F_A)$ je automat, který rozeznává L, a nechť $\mathcal{T} = (Q_T, \Sigma, q_0^T, \delta_T, \Delta, \lambda)$. Zkonstruujeme deterministický automat $\mathcal{B}$, který rozeznává $\mathcal{T}^{-1}(L)$. Na vstupu $u \in \Sigma^*$ chceme simulovat překlad u na $\mathcal{T}(u)$ a současně přijímání slova $\mathcal{T}(u)$ automatem $\mathcal{A}$.

Množina stavů $\mathcal{B}$ je $Q_T \times Q_A$ a jeho přechodová funkce je definována předpisem

$$(q^T, q^A) \cdot c = (q^T \cdot c, q^A \cdot \lambda(q^T, c)).$$



Množina přijímajících stavů je $Q_T \times F_A$. Z definice přechodové funkce plyne, že

$$(q_0^T, q_0^A) \cdot u = \left(q_0^T \cdot u, q_0^A \cdot \mathcal{T}(u)\right).$$

Automat $\mathcal{B}$ tedy přijímá u, právě když $\mathcal{A}$ přijímá $\mathcal{T}(u)$, což jsme chtěli. $\square$

Věta. Je-li $L \subseteq \Sigma^*$ regulární jazyk, je i $\mathcal{T}(L) \subseteq \Delta^*$ regulární.

Důkaz. Nechť je $\mathcal{A} = (Q_A, \Sigma, q_0^A, \delta_A, F_A)$ deterministický automat, který rozeznává L, a nechť opět $\mathcal{T} = (Q_T, \Sigma, q_0^T, \delta_T, \Delta, \lambda)$. Slovo w leží v $\mathcal{T}(L)$, právě když existuje slovo $u = c_1 c_2 \cdots c_k \in L$ takové, že $\mathcal{T}(u) = w$. Slovo u definuje faktorizaci

$$w = w_1 w_2 \cdots w_k,$$

kde pro každé $1 \leq i < k$ platí

$$\lambda(q_{i-1}^T, c_i) = w_i, \quad \text{kde} \quad q_{i-1}^T = q_0^T \cdot c_1 c_2 \cdots c_{i-1}.$$

Sestrojíme nedeterministický automat $\mathcal{D}$, který bude takové faktorizace přijímat.

Stavy automatu $\mathcal{D}$ budou prvky $Q_T \times Q_A \times W$, kde W je množina slov, která jsou sufixem nějakého obrazu funkce λ. Automat $\mathcal{D}$ na vstupu w „hádá“ možnou faktorizaci $w = w_1 w_2 \cdots w_k$. Nechť je již zpracován nějaký prefix $w_1 w_2 \cdots w_i$, který odpovídá slovu $u' = c_1 c_2 \cdots c_i$, tedy $w_1 w_2 \cdots w_i = \mathcal{T}(u')$. Nechť $q^T = q_0^T \cdot u'$ a $q^A = q_0^A \cdot u'$ a nechť $w = w_1 w_2 \cdots w_i w'$. Nyní je třeba vybrat písmeno c_{i+1} takové, že $w_{i+1} = \lambda(q^T, c_{i+1})$ je prefixem w', přečíst w_{i+1} a změnit stav q^T na $q^T \cdot c_{i+1}$ a stav q^A na $q^A \cdot c_{i+1}$. Opět tedy chceme automatem $\mathcal{D}$ simulovat pomocí direktního součinu běh překladače i automatu. Komplikace spočívá v tom, že slovo w_{i+1} nemusí být jen jedno písmeno, a jeho přečtení tedy sestává z několika kroků. Nechť je $w_{i+1} = \ell_1 \ell_2 \cdots \ell_m$. Stav $(q^T \cdot c_{i+1}, q^A \cdot c_{i+1}, \ell_{s+1} \ell_{s+2} \cdots \ell_m)$ automatu $\mathcal{D}$ znamená, že na cestě do stavu $(q^T \cdot c_{i+1}, q^A \cdot c_{i+1})$, který zapisujeme jako $(q^T \cdot c_{i+1}, q^A \cdot c_{i+1}, \varepsilon)$, zbývá přečíst slovo $\ell_{s+1} \ell_{s+2} \cdots \ell_m$.

Přechodová relace (nedeterministického) automatu $\mathcal{D}$ tedy obsahuje následující tři typy přechodů

$$\text{(i)} \qquad (q^T,\, q^A,\, \varepsilon) \xrightarrow{\ \ell\ } (q^T \cdot c,\, q^A \cdot c,\, w'),$$

kde c je libovolné písmeno ze Σ, $\ell \in \Delta$ a $\ell w' = \lambda(q^T, c)$. Tento přechod představuje volbu písmene c a přečtení prvního písmene z příslušného $w = \ell w'$. Přečtení dalších písmen (včetně posledního) je dáno přechody typu

$$\text{(ii)} \qquad (q^T,\, q^A,\, \ell w') \xrightarrow{\ \ell\ } (q^T,\, q^A,\, w').$$

Zbývá ošetřit speciální situaci, kdy $\lambda(q^T, c)$ je prázdné slovo. Tomu odpovídá ε-přechod

$$\text{(iii)} \qquad (q^T,\, q^A,\, \varepsilon) \xrightarrow{\ \varepsilon\ } (q^T \cdot c,\, q^A \cdot c,\, \varepsilon).$$

Přijímající stavy jsou $Q_T \times F_A \times \{\varepsilon\}$. Z konstrukce plyne, že pokud $w = \mathcal{T}(u)$, $u \in \Sigma$, pak existuje automatem $\mathcal{D}$ cesta z $(q_0^T, q_0^A, \varepsilon)$ do stavu $(q_0^T \cdot u, q_0^A \cdot u, \varepsilon)$ a pokud $u \in L$, je slovo w přijato. Naopak, je-li w přijato stavem $(q_f^T, q_f^A, \varepsilon)$, můžeme přijímající cestu grafem rozdělit na úseky procházející stavy tvaru (q^T, q^A, ε) a najít tak slovo $u \in L$, pro které je $q_0^A \cdot u = q_f^A$ a $\mathcal{T}(u) = w$. □