Substitúcie

Peter Kostolányi

21. februára 2017

Definícia

Pod substitúciou sa v teórii formálnych jazykov chápe „zovšeobecnenie“ homomorfizmu, kde obrazom slov nie sú slová, ale jazyky.

Definícia 1. Nech Σ, Γ sú abecedy a $\tau \colon \Sigma^* \to 2^{\Gamma^*}$ je zobrazenie. Zobrazenie τ sa nazýva *substitúcia*, ak $\tau(\varepsilon) = \{\varepsilon\}$ a pre všetky $u, v \in \Sigma^*$ platí

$$\tau(uv) = \tau(u)\tau(v).$$

Poznámka 1. Požiadavka $\tau(\varepsilon) = \{\varepsilon\}$ je v uvedenej definícii skutočne podstatná, pretože *nevyplýva* z vlastnosti $\tau(uv) = \tau(u)\tau(v)$ pre všetky $u, v \in \Sigma^*$, čím sa substitúcie odlišujú od homomorfizmov. Nájdenie vhodného protipríkladu je jednou z úloh určených na nasledujúce cvičenie.

Poznámka 2. Ak $h \colon \Sigma^* \to \Gamma^*$ je homomorfizmus, zobrazenie $\tau \colon \Sigma^* \to 2^{\Gamma^*}$ definované pre všetky $w \in \Sigma^*$ ako $\tau(w) = \{h(w)\}$ je zjavne substitúcia. Aj keď teda homomorfizmus z formálneho hľadiska nie je substitúciou, možno sa na substitúcie z určitého pohľadu dívať ako na zovšeobecnenie homomorfizmov.

Poznámka 3. Definíciu substitúcie možno interpretovať aj tak, že ide o homomorfizmus monoidov $(\Sigma^*, \cdot)$ a $(2^{\Gamma^*}, \cdot)$.

Z nasledujúceho tvrdenia vyplýva, že podobne ako homomorfizmy sú aj substitúcie jednoznačne určené obrazmi jednotlivých písmen, a teda ich možno alternatívne definovať aj ako zobrazenia zo Σ do 2^{Γ^*} (s následným rozšírením na Σ^*). V nasledujúcom budeme túto skutočnosť využívať bez toho, aby sme na to explicitne upozorňovali.

Tvrdenie 1. *Nech Σ, Γ sú abecedy a $\tau \colon \Sigma^* \to 2^{\Gamma^*}$ je substitúcia. Nech $\tau' \colon \Sigma^* \to 2^{\Gamma^*}$ je substitúcia taká, že pre všetky $c \in \Sigma$ platí $\tau'(c) = \tau(c)$. Potom $\tau' = \tau$.*

Dôkaz. Treba ukázať, že pre všetky $w \in \Sigma^*$ platí $\tau'(w) = \tau(w)$. To možno urobiť jednoduchou indukciou vzhľadom na dĺžku slova w (detaily prenechávame čitateľovi). $\square$

Podobne ako v prípade homomorfizmov možno definíciu substitúcie aditívne rozšíriť aj na jazyky.

Definícia 2. Nech Σ, Γ sú abecedy, $\tau \colon \Sigma^* \to 2^{\Gamma^*}$ je substitúcia a $L \subseteq \Sigma^*$ je jazyk. *Obraz jazyka L pri zobrazení substitúciou τ* je jazyk

$$\tau(L) = \bigcup_{w \in L} \tau(w).$$

Príklad 1. Nech $\Sigma = \{a, b\}$ a substitúcia $\tau \colon \Sigma^* \to 2^{\Sigma^*}$ je daná ako $\tau(a) = \{\varepsilon, b\}$ a $\tau(b) = \{a\}^*$. Potom pre $L = \{ab, bb\}$ platí

$$\tau(L) = \tau(ab) \cup \tau(bb) = \big(\{a\}^* \cup \{b\}\{a\}^*\big) \cup \{a\}^* = \{a\}^* \cup \{b\}\{a\}^*.$$

Je zrejmé, že substitúcia je veľmi silná operácia: stačí si uvedomiť, že jazyk $\tau(a)$ napríklad nemusí byť ani rekurzívne vyčísliteľný. Z tohto dôvodu sa v súvislosti s konkrétnymi triedami jazykov budeme zaoberať väčšinou substitúciami, ktorých obor hodnôt je určitým spôsobom obmedzený. To odôvodňuje zavedenie pojmu $\mathscr{L}$-substitúcie pre triedu jazykov $\mathscr{L}$.

Definícia 3. Nech Σ, Γ sú abecedy, $\tau \colon \Sigma^* \to 2^{\Gamma^*}$ je substitúcia a $\mathscr{L}$ je trieda jazykov. Hovoríme, že τ je $\mathscr{L}$-substitúcia, ak pre všetky $c \in \Sigma$ platí $\tau(c) \in \mathscr{L}$.

Poznámka 4. Treba upozorniť na skutočnosť, že z uvedenej definície ešte pre $\mathscr{L}$-substitúciu $\tau\colon \Sigma^* \to 2^{\Gamma^*}$ vo všeobecnosti *nevyplýva* $\tau(w) \in \mathscr{L}$ pre všetky $w \in \Sigma^*$. Túto vlastnosť však zjavne majú všetky $\mathscr{L}$-substitúcie pre triedy $\mathscr{L}$ uzavreté na zreťazenie (čo sú napr. všetky triedy jazykov Chomského hierarchie).

V podobnom duchu tiež definujeme *regulárnu substitúciu* ako $\mathscr{R}$-substitúciu, *bezkontextovú substitúciu* ako $\mathscr{L}_{CF}$-substitúciu a podobnú terminológiu budeme používať aj pre ďalšie známe triedy jazykov.

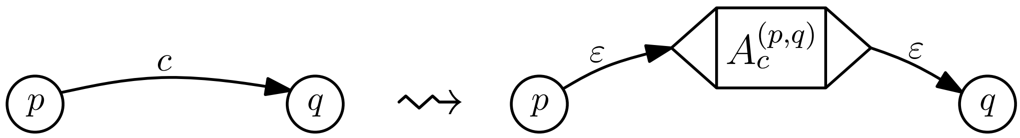
Uzavretosť tried $\mathscr{R}$ a $\mathscr{L}_{CF}$ na substitúciu

Z predchádzajúcich úvah okrem iného vyplýva, že nemá zmysel očakávať uzavretosť nejakej zmysluplnej a netriviálnej triedy jazykov na všeobecnú substitúciu. Aj preto budeme hovoriť, že trieda $\mathscr{L}$ *je uzavretá na substitúciu*, ak je uzavretá na $\mathscr{L}$-substitúciu. V nasledujúcom dokážeme, že triedy $\mathscr{R}$ a $\mathscr{L}_{CF}$ sú uzavreté na substitúciu.

Veta 1. *Trieda $\mathscr{R}$ je uzavretá na substitúciu.*

Dôkaz. Nech $L \subseteq \Sigma^*$ je regulárny jazyk akceptovaný deterministickým konečným automatom $A = (K, \Sigma, \delta, q_0, F)$ a $\tau\colon \Sigma^* \to 2^{\Gamma^*}$ je regulárna substitúcia. Zostrojíme nedeterministický konečný automat A' taký, že $L(A') = \tau(L)$.

Keďže je substitúcia τ regulárna, pre každé $c \in \Sigma$ existuje nedeterministický konečný automat A_c v „prasiatkovom“ normálnom tvare taký, že $L(A_c) = \tau(c)$. Automat A' zostrojíme z automatu A tak, že každý prechod na symbol c nahradíme samostatnou kópiou automatu A_c tak, ako je znázornené na obrázku 1. Každá kópia automatu A_c musí byť označená počiatočným a koncovým stavom prechodu, ktorý nahrádza – inak by mohlo dôjsť k „pomiešaniu stavov“ jednotlivých kópií automatu A_c.



Obr. 1: Každý prechod automatu A na písmeno c nahradíme samostatnou kópiou automatu A_c.

Predpokladajme, že pre všetky $c \in \Sigma$ je $A_c = (K[A_c], \Gamma, \delta[A_c], q_0[A_c], \{q_F[A_c]\})$, pričom pre všetky $a, b \in \Sigma$ také, že $a \neq b$ sú množiny stavov $K[A_a]$ a $K[A_b]$ disjunktné. Formálnu konštrukciu automatu A' potom môžeme zapísať nasledovne: $A' = (K', \Sigma', \delta', q_0', F')$, kde

$$K' = K \cup \bigcup_{c \in \Sigma} \left(K[A_c] \times K^2\right),$$

$\Sigma' = \Gamma$, $q_0' = q_0$, $F' = F$ a kde prechodová funkcia δ' je daná nasledovne: pre všetky stavy $p \in K$ definujeme

$$\delta'(p, \varepsilon) = \{(q_0[A_c], p, \delta(p, c)) \mid c \in \Sigma\},$$

čo zodpovedá prechodom na ε vedúcim zo stavov z množiny K do počiatočných stavov kópií automatov A_c. Ďalej pre všetky $p, r \in K$, $c \in \Sigma$ a pre všetky $q \in K[A_c] - \{q_F[A_c]\}$ a $a \in \Gamma \cup \{\varepsilon\}$ definujeme prechod

$$\delta'((q, p, r), a) = (\delta[A_c](q, a), p, r)$$

v rámci zodpovedajúcej kópie automatu A_c a nakoniec, pre všetky $p, r \in K$ a $c \in \Sigma$ definujeme prechod

$$\delta'((q_F[A_c], p, r), \varepsilon) = r$$

vedúci z akceptačného stavu danej kópie automatu A_c do zodpovedajúceho stavu z K. Konštrukcia je hotová. Dôkaz rovnosti $L(A') = \tau(L)$ robiť nebudeme. $\square$

Veta 2. *Trieda $\mathscr{L}_{CF}$ je uzavretá na substitúciu.*

Dôkaz. Nech L je bezkontextový jazyk generovaný bezkontextovou gramatikou $G = (N, T, P, \sigma)$. Nech τ je bezkontextová substitúcia na T^*. Gramatiku G' generujúcu jazyk $L(G') = \tau(L)$ možno získať nahradením každého terminálu $c \in T$ počiatočným neterminálom bezkontextovej gramatiky pre $\tau(c)$ (za predpokladu disjunktnosti jednotlivých množín neterminálov a neexistencie symbolu, ktorý je súčasne neterminálom jednej gramatiky a terminálom inej gramatiky).

Presnejšie: nech pre každé $c \in T$ je $G_c = (N_c, T_c, P_c, \sigma_c)$ bezkontextová gramatika taká, že $L(G_c) = \tau(c)$. Predpokladajme navyše, že platí

$$\begin{aligned}
N_c \cap N_d = \emptyset \qquad & \forall c, d \in T, c \neq d, \\
N_c \cap N = \emptyset \qquad & \forall c \in T, \\
\left(N \cup \bigcup_{c \in T} N_c \right) \cap \left(\bigcup_{c \in T} T_c \right) = \emptyset. &
\end{aligned}$$

Uvažujme homomorfizmus h na $(N \cup T)^*$ taký, že pre všetky $\xi \in N$ je $h(\xi) = \xi$ a pre všetky $c \in T$ je $h(c) = \sigma_c$. Gramatiku G' potom môžeme zostrojiť nasledovne: $G' = (N', T', P', \sigma')$, kde

$$\begin{aligned}
N' &= N \cup \bigcup_{c \in T} N_c, \\
T' &= \bigcup_{c \in T} T_c, \\
P' &= \{\xi \to h(w) \mid (\xi \to w) \in P\} \cup \bigcup_{c \in T} P_c, \\
\sigma' &= \sigma.
\end{aligned}$$

Konštrukcia je hotová (pričom dôkaz rovnosti $L(G') = \tau(L)$ vynechávame). $\square$