

# Myhillova-Nerodova veta

Peter Kostolányi

28. februára 2017

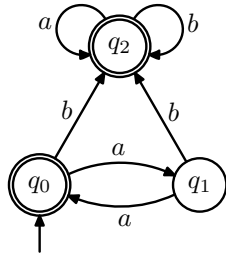
## Deterministické konečné automaty a relácie ekvivalencie na $\Sigma^*$

Nech  $A = (K, \Sigma, \delta, q_0, F)$  je *deterministický* konečný automat (s úplnou prechodovou funkciou). Automat  $A$  prirodzeným spôsobom definuje reláciu  $R_A$  na  $\Sigma^*$  takú, že pre dvojicu slov  $u, v \in \Sigma^*$  platí  $uR_Av$  práve vtedy, keď automat  $A$  dočíta slová  $u$  a  $v$  v rovnakom stave. Formálne:

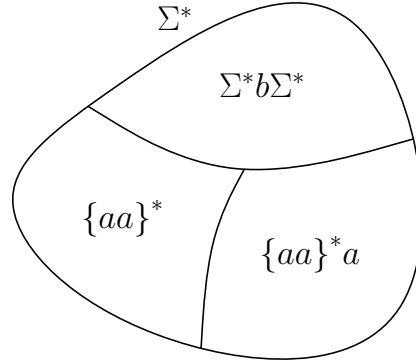
$$\forall u, v \in \Sigma^* : uR_Av \iff (\exists q \in K : (q_0, u) \vdash_A^* (q, \varepsilon) \wedge (q_0, v) \vdash_A^* (q, \varepsilon)).$$

Je triviálnou úlohou overiť, že takto definovaná relácia  $R_A$  je *reflexívna* (automat  $A$  slová  $w$  a  $w$  vždy dočíta v rovnakom stave), *symetrická* (ak  $A$  dočíta slová  $u, v$  v rovnakom stave, tak dočíta v rovnakom stave aj slová  $v, u$ ) a *tranzitívna* (ak  $A$  dočíta v rovnakom stave slová  $x, y$  a  $y, z$ , tak dočíta v rovnakom stave aj slová  $x, z$ ). Relácia  $R_A$  je teda *reláciou ekvivalencie* na  $\Sigma^*$ .

Príklad deterministického konečného automatu  $A$  nad abecedou  $\Sigma = \{a, b\}$  a jemu zodpovedajúcej relácie  $R_A$  (určujúcej rozklad  $\Sigma^*$  na tri triedy ekvivalencie) je na obrázku 1.



(a) Deterministický konečný automat  $A$ .



(b) Triedy ekvivalencie relácie  $R_A$ .

**Obr. 1:** Deterministický konečný automat  $A$  a jemu zodpovedajúca relácia ekvivalencie  $R_A$  na  $\Sigma^*$ .

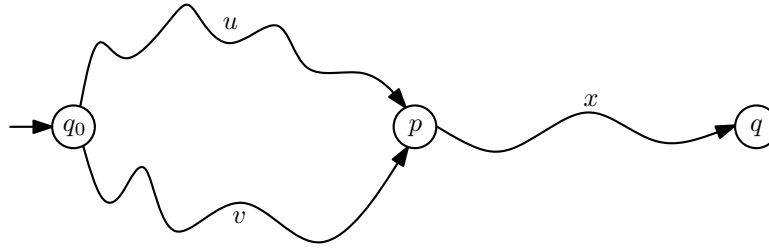
Okamžite si možno všimnúť tri podstatné vlastnosti relácie ekvivalencie  $R_A$ :

1. Nech  $u, v \in \Sigma^*$  sú slová také, že  $uR_Av$ . Potom platí buď  $u, v \in L(A)$ , alebo  $u, v \notin L(A)$ . Slová  $u$  a  $v$  totiž automat  $A$  dočíta v rovnakom stave, ktorý je buď akceptačný, alebo neakceptačný. V každej triede ekvivalencie relácie  $R_A$  tak sú buď výlučne slová z  $L(A)$ , alebo výlučne slová z komplementu  $L(A)$ . Jazyk  $L(A)$  je preto *zjednotením niekoľkých tried ekvivalencie relácie  $R_A$* . Napríklad pre automat  $A$  na obrázku 1 platí  $L(A) = \Sigma^*b\Sigma^* \cup \{aa\}^*$ .
2. Keďže má automat  $A$  konečne veľa stavov, relácia  $R_A$  má konečne veľa tried ekvivalencie. Hovoríme, že relácia  $R_A$  je *konečného indexu* a počet jej tried ekvivalencie nazývame *indexom* relácie  $R_A$ .
3. Ak automat  $A$  dočíta slová  $u, v$  v rovnakom stave, dočíta v rovnakom stave aj slová  $ux, vx$ , kde  $x$  je ľubovoľné – táto situácia je schematicky znázornená na obrázku 2. Inými slovami, relácia  $R_A$  je *sprava invariantná*, čiže pre všetky  $u, v \in \Sigma^*$  platí

$$uR_Av \Rightarrow \forall x \in \Sigma^* : uxR_Avx.$$

Priamym dôsledkom uvedených troch pozorovaní je nasledujúce tvrdenie.

**Tvrdenie 1.** *Nech  $L \subseteq \Sigma^*$  je regulárny jazyk. Potom na  $\Sigma^*$  existuje sprava invariantná relácia ekvivalencie  $R$  konečného indexu taká, že  $L$  je zjednotením niekoľkých jej tried ekvivalencie.*



**Obr. 2:** Ak deterministický konečný automat  $A$  dočíta slová  $u$  a  $v$  v rovnakom stave, dočíta v rovnakom stave aj slová  $ux$  a  $vx$  pre ľubovoľné slovo  $x$ .

Uvažujme teraz ľubovoľný jazyk  $L \subseteq \Sigma^*$ , ktorý je zjednotením niekoľkých tried ekvivalencie nejakej sprava invariantnej relácie ekvivalencie konečného indexu  $R$ . Ukážeme, že relácia  $R$  spoločne s informáciou, ktoré triedy ekvivalencie zodpovedajú slovám z  $L$ , jednoznačne určuje deterministický konečný automat  $A = (K, \Sigma, \delta, q_0, F)$  akceptujúci  $L$ .

Označme symbolom  $[w]_R$  triedu ekvivalencie relácie  $R$  obsahujúcu slovo  $w$ . Automat  $A$  potom skonštruujeme nasledovne:

- Za množinu stavov  $K$  vezmeme množinu tried ekvivalencie relácie  $R$ , čo možno formálne zapísať ako  $K = \{[w]_R \mid w \in \Sigma^*\}$ . Táto množina je konečná, pretože relácia  $R$  je konečného indexu.
- Pre každú triedu ekvivalencie  $[u]_R$  (kde  $u \in \Sigma^*$ ) a každé  $c \in \Sigma$  položíme  $\delta([u]_R, c) = [uc]_R$ . Takto daná prechodová funkcia je korektná iba v prípade, že jej výstup *nezávisí na voľbe reprezentanta* triedy  $[u]_R$ . Treba teda ukázať, že ak platí  $[u]_R = [v]_R$  (čiže ak  $uRv$ ), tak nutne platí aj  $[uc]_R = [vc]_R$ . Keďže je ale relácia  $R$  sprava invariantná,  $uRv$  implikuje  $ucRvc$ , a teda aj  $[uc]_R = [vc]_R$ .
- Za počiatočný stav vezmeme triedu ekvivalencie obsahujúcu prázdne slovo:  $q_0 = [\varepsilon]_R$ .
- Za množinu akceptačných stavov vezmeme množinu tých tried ekvivalencie relácie  $R$ , ktoré obsahujú slová z  $L$ ; čiže  $F = \{[w]_R \mid w \in L\}$ . Korektnosť uvedeného zápisu je dôsledkom skutočnosti, že  $L$  je zjednotením niekoľkých tried ekvivalencie relácie  $R$  – v každej triede ekvivalencie sú teda výlučne slová z  $L$  alebo výlučne slová z  $L^C$ .

Tvrdenie  $L(A) = L$  dokazovať nebudeme, ale jeho platnosť by mala byť očividná. Dôsledkom je obrátená implikácia z tvrdenia 1:

**Tvrdenie 2.** *Nech  $L \subseteq \Sigma^*$  je zjednotením niekoľkých tried ekvivalencie nejakej sprava invariantnej relácie ekvivalencie konečného indexu na  $\Sigma^*$ . Potom je jazyk  $L$  regulárny.*

Ľahko možno overiť, že deterministický konečný automat zostrojený pomocou uvedenej konštrukcie k relácii ekvivalencie na obrázku 1b a k jazyku  $L = \Sigma^*b\Sigma^* \cup \{aa\}^*$  je až na izomorfizmus (teda premenovanie stavov) zhodný s konečným automatom na obrázku 1a. Čitateľa tak snáď netreba presviedčať o skutočnosti, že *deterministické konečné automaty a sprava invariantné relácie ekvivalencie konečného indexu sú iba „odlišnými pohľadmi na ten istý objekt“*. Toto spojenie je pre teóriu okolo Myhillovej-Nerodovej vety kľúčové.

### Pravá syntaktická ekvivalencia

V nasledujúcom ukážeme, že k ľubovoľnému (nie nutne regulárnemu) jazyku  $L \subseteq \Sigma^*$  možno definovať kanonickú sprava invariantnú reláciu ekvivalencie na  $\Sigma^*$  – tzv. *pravú syntaktickú ekvivalenciu* – takú, že  $L$  je zjednotením časti jej tried. Táto relácia je navyše konečného indexu práve vtedy, keď je jazyk  $L$  regulárny; počet jej tried ekvivalencie je v takom prípade rovný *najmenšiemu možnému* počtu stavov deterministického konečného automatu akceptujúceho  $L$ .

**Definícia 1.** Nech  $L \subseteq \Sigma^*$  je jazyk. *Pravá syntaktická ekvivalencia* indukovaná jazykom  $L$  je relácia  $R_L$  taká, že pre všetky  $u, v \in \Sigma^*$  platí

$$uR_Lv \iff (\forall z \in \Sigma^* : uz \in L \iff vz \in L).$$

**Tvrdenie 3.** Nech  $L \subseteq \Sigma^*$  je jazyk. *Pravá syntaktická ekvivalencia  $R_L$  je sprava invariantná relácia ekvivalencie na  $\Sigma^*$  a jazyk  $L$  je zjednotením časti jej tried.*

*Dôkaz.* Dôkaz, že  $R_L$  je relácia ekvivalencie, je triviálny a prenechaný čitateľovi. Ak  $uR_Lv$  a ak zvolíme  $z = \varepsilon$ , dostávame  $u \in L$  práve vtedy, keď  $v \in L$ ; inými slovami,  $L$  je zjednotením časti tried ekvivalencie relácie  $R_L$ . Dokážeme, že relácia  $R_L$  je sprava invariantná. Nech  $u, v \in \Sigma^*$  sú slová také, že  $uR_Lv$  a  $x \in \Sigma^*$  je ľubovoľné. Z definície  $R_L$  vyplýva, že pre všetky  $z \in \Sigma^*$  platí  $uz \in L \iff vz \in L$ . Špeciálne táto ekvivalencia platí pre všetky  $z$  také, že  $z = xy$  pre nejaké  $y \in \Sigma^*$ . Inými slovami, pre všetky  $y \in \Sigma^*$  platí  $uxy \in L \iff vxy \in L$ , z čoho podľa definície relácie  $R_L$  vyplýva  $uxR_Lvx$ . Relácia  $R_L$  teda je sprava invariantná.  $\square$

Význam pravej syntaktickej ekvivalencie spočíva predovšetkým v skutočnosti, že ide, ako dokážeme v tvrdení 4, o *najhrubšiu sprava invariantnú reláciu ekvivalencie na  $\Sigma^*$  takú, že  $L$  je zjednotením časti jej tried*. To znamená, že ak je jazyk  $L$  zjednotením časti tried ekvivalencie nejakej sprava invariantnej relácie ekvivalencie  $R$ , tak  $R$  vznikne z relácie  $R_L$  „rozbitím“ jej tried ekvivalencie (teda  $R$  je *zjemnením* relácie  $R_L$ ).

**Tvrdenie 4.** Nech  $L \subseteq \Sigma^*$  je jazyk a  $R$  je sprava invariantná relácia ekvivalencie na  $\Sigma^*$  taká, že  $L$  je zjednotením časti jej tried. Nech  $u, v \in \Sigma^*$  sú slová také, že  $uRv$ . Potom  $uR_Lv$ .

*Dôkaz.* Z pravej invariance relácie  $R$  vyplýva, že pre všetky  $x \in \Sigma^*$  platí  $uxRvx$ . Keďže je ale jazyk  $L$  zjednotením časti tried relácie  $R$ , v takom prípade platí buď  $ux, vx \in L$ , alebo  $ux, vx \notin L$ . Inými slovami, pre všetky  $x \in \Sigma^*$  je  $ux \in L \iff vx \in L$ , a teda  $uR_Lv$ .  $\square$

Ako sme dokázali vyššie, jazyk  $L \subseteq \Sigma^*$  je regulárny práve vtedy, keď na  $\Sigma^*$  existuje sprava invariantná relácia ekvivalencie *konečného indexu* taká, že  $L$  je zjednotením niekoľkých jej tried. Je zrejmé, že ak nie je konečného indexu najhrubšia takáto relácia, tak nemôže byť konečného indexu ani žiadna iná. Preto dostávame nasledujúci dôsledok.

**Dôsledok 1.** Nech  $L \subseteq \Sigma^*$  je jazyk. *Jazyk  $L$  je regulárny práve vtedy, keď je relácia  $R_L$  konečného indexu.*

V prípade, že je jazyk  $L$  regulárny, má relácia  $R_L$  *najmenší počet tried ekvivalencie* spomedzi všetkých sprava invariantných relácií ekvivalencie konečného indexu  $R$  takých, že  $L$  je zjednotením niekoľkých tried relácie  $R$  – všetky takéto relácie sú totiž zjemnením relácie  $R_L$ . Ľahko tiež vidieť, že  $R_L$  je *jediná* relácia na  $\Sigma^*$  s touto vlastnosťou; jediným zjemnením relácie  $R_L$  s rovnakým počtom tried je totiž sama relácia  $R_L$ .

Ak sa (v zmysle predchádzajúceho oddielu) na toto tvrdenie pozrieme z perspektívy konečných automatov, zistíme, že relácia  $R_L$  zodpovedá deterministickému konečnému automatu s najmenším počtom stavov, ktorý akceptuje jazyk  $L$ . Tento automat je navyše (až na izomorfizmus, t.j. premenovanie stavov) určený jednoznačne a nazývame ho *minimálnym automatom* pre jazyk  $L$ .

**Dôsledok 2.** Nech  $L \subseteq \Sigma^*$  je regulárny jazyk. *Potom existuje (až na izomorfizmus) jednoznačne určený deterministický konečný automat  $A = (K, \Sigma, \delta, q_0, F)$  taký, že  $L(A) = L$  a súčasne pre každý deterministický konečný automat  $A' = (K', \Sigma', \delta', q'_0, F')$  s  $L(A') = L$  platí  $|K'| \geq |K|$ . Automat  $A$  sa nazýva *minimálny automat pre jazyk  $L$*  a platí  $R_A = R_L$ .*

Ako hlavné posolstvo tohto a predchádzajúceho oddielu ešte raz zdôraznime skutočnosť, že *existuje prirodzená korešpondencia medzi deterministickými konečnými automaty nad abecedou  $\Sigma$  a sprava invariantnými reláciami ekvivalencie konečného indexu na  $\Sigma^*$* . Navyše, špeciálny prípad takejto relácie – *pravá syntaktická ekvivalencia indukovaná jazykom  $L$*  – zodpovedá špeciálnemu deterministickému konečnému automatu – *minimálnemu automatu pre jazyk  $L$* .

## Myhillova-Nerodova veta

Tvrdenie známe ako *Myhillova-Nerodova veta* (aj keď pri tomto jej znení by možno bolo presnejšie iba „Nerodova veta“) sme už v podstate dokázali v predchádzajúcich dvoch oddieloch.

**Veta 1** (Myhill, Nerode). *Nech  $L \subseteq \Sigma^*$  je jazyk. Nasledujúce tvrdenia sú ekvivalentné:*

- (i)  *$L$  je regulárny.*
- (ii) *Existuje sprava invariantná relácia ekvivalencie konečného indexu na  $\Sigma^*$  taká, že  $L$  je zjednotením niekoľkých jej tried.*
- (iii) *Pravá syntaktická ekvivalencia  $R_L$  je konečného indexu.*

*Dôkaz.* Vyplýva priamo z tvrdenia 1, tvrdenia 2 a dôsledku 1. □

## Použitie Myhillovej-Nerodovej vety

V rámci tohto oddielu ukážeme dva typické príklady použitia Myhillovej-Nerodovej vety. Prvým z nich je konštrukcia minimálneho automatu k danému regulárnemu jazyku. Postup, ktorý využijeme pri riešení nasledujúcej ukázkovej úlohy pozostáva z nájdenia pravej syntaktickej ekvivalencie  $R_L$  pre jazyk  $L$  a z dôkazu, že skutočne ide o pravú syntaktickú ekvivalenciu.

**Úloha 1.** Nech  $L = \{w \in \{a, b\}^* \mid \#_a(w) \equiv 7 \pmod{11} \wedge \#_b(w) \equiv 19 \pmod{23}\}$ . Nájdite minimálny (deterministický konečný) automat pre jazyk  $L$ .

*Riešenie.* Definujme na  $\Sigma^* = \{a, b\}^*$  reláciu  $R$  nasledovne:

$$\forall u, v \in \Sigma^* : uRv \iff \#_a(u) \equiv \#_a(v) \pmod{11} \wedge \#_b(u) \equiv \#_b(v) \pmod{23}.$$

Ľahko možno overiť, že  $R$  je sprava invariantnou reláciou ekvivalencie na  $\Sigma^*$  s  $11 \cdot 23 = 253$  triedami ekvivalencie ( $R$  je teda konečného indexu). Navyše  $L = [a^7 b^{19}]_R$ , a teda  $L$  je zjednotením niekoľkých tried ekvivalencie relácie  $R$  (presnejšie jednej). Poriadny dôkaz uvedených tvrdení prenechávame čitateľovi (ktorému je silno odporúčané chopiť sa tejto príležitosti).

Relácia  $R$  je zjemnením  $R_L$  – ak teda pre nejaké  $u, v \in \Sigma^*$  platí  $uRv$ , tak aj  $uR_L v$ . Na dôkaz  $R = R_L$  teda stačí ukázať, že ak pre nejaké  $u, v \in \Sigma^*$  platí  $uR_L v$ , tak nutne aj  $uRv$ .

Nepriamo. Nech neplatí  $uRv$ , čiže  $\#_a(u) \not\equiv \#_a(v) \pmod{11}$  alebo  $\#_b(u) \not\equiv \#_b(v) \pmod{23}$ . Ak teraz vezmeme slovo  $z \in \Sigma^*$  také, že  $\#_a(uz) \equiv 7 \pmod{11}$  a  $\#_b(uz) \equiv 19 \pmod{23}$ , tak táto vlastnosť nemôže platiť pre slovo  $vz$ . Teda  $uz \in L$ , kým  $vz \notin L$ . Preto neplatí ani  $uR_L v$ .

Dokázali sme teda, že skutočne  $R = R_L$ . Na základe prirodzenej korešpondencie medzi sprava invariantnými reláciami ekvivalencie konečného indexu a deterministickými konečnými automatmi potom môžeme uzavrieť, že minimálny automat pre jazyk  $L$  je daný ako  $A = (K, \Sigma, \delta, q_0, F)$ , kde  $K = \mathbb{Z}_{11} \times \mathbb{Z}_{23}$ ,  $\Sigma = \{a, b\}$ ,  $q_0 = [0, 0]$ ,  $F = \{[7, 19]\}$  a prechodová funkcia  $\delta$  je pre každé  $[i, j] \in K$  daná predpismi  $\delta([i, j], a) = [i + 1, j]$  a  $\delta([i, j], b) = [i, j + 1]$ . □

Myhillovu-Nerodovu vetu možno využiť aj ako efektívny nástroj na vyvracanie regularity jazykov, ako ukážeme na nasledujúcej úlohe.

**Úloha 2.** Dokážte, že jazyk  $L = \{a^{n^2} \mid n \in \mathbb{N}\}$  nie je regulárny.

*Riešenie.* Dokážeme, že pravá syntaktická ekvivalencia  $R_L$  indukovaná jazykom  $L$  nie je konečného indexu, z čoho podľa Myhillovej-Nerodovej vety vyplýva, že  $L$  nie je regulárny.

Zrejme stačí ukázať, že (pre ľubovoľné  $n, m \in \mathbb{N}$ ) slovo  $a^{n^2}$  nemôže byť v relácii  $R_L$  so slovom  $a^{m^2}$  takým, že  $m > n$ . Za účelom sporu predpokladajme opak a vezmime  $z = a^{2n+1}$ . Potom  $a^{n^2} z = a^{n^2+2n+1} = a^{(n+1)^2} \in L$ , ale  $a^{m^2} z = a^{m^2+2n+1} \notin L$ , keďže  $(m+1)^2 = m^2 + 2m + 1 > m^2 + 2n + 1$ . To je spor s definíciou relácie  $R_L$ . □

Je jedným z kľúčových výsledkov teórie automatov, že neexistuje žiadne  $N \in \mathbb{N}$  také, že všetky regulárne jazyky možno akceptovať deterministickým konečným automatom s najviac  $N$  stavmi. Cieľom nasledujúcej úlohy je dokázať toto tvrdenie.

**Úloha 3.** Dokážte, že ku každému  $n \in \mathbb{N}$  existuje regulárny jazyk  $L_n \in \mathcal{R}$  taký, že ľubovoľný deterministický konečný automat akceptujúci  $L_n$  musí mať aspoň  $n$  stavov.

*Riešenie.* Tvrdenie očividne stačí dokázať pre  $n \geq 2$ . K ľubovoľnému takému  $n$  vezmeme jazyk

$$L_n = \{a^k \mid k \in \mathbb{N}; k \equiv 0 \pmod{n}\}.$$

Dokážeme, že ľubovoľný deterministický konečný automat akceptujúci  $L_n$  musí mať aspoň  $n$  stavov.

Uvažujme pravú syntaktickú ekvivalenciu  $R_{L_n}$ . Nech  $i, j \in \{0, \dots, n-1\}$  a  $i \neq j$ . Ak vezmeme  $z = a^{n-i}$ , tak potom zjavne  $a^i z \in L_n$ , kým  $a^j z \notin L_n$ . Slová  $a^i$  a  $a^j$  teda nemôžu byť v rovnakej triede pravej syntaktickej ekvivalencie  $R_{L_n}$ , ktorá tak musí mať aspoň  $n$  tried. Dôsledkom je, že deterministický konečný automat akceptujúci  $L_n$  musí mať aspoň  $n$  stavov.  $\square$