Cvičenie č. 5
Konečné automaty

Peter Kostolányi

23. októbra 2024

1 Deterministické a nedeterministické konečné automaty

Začnime stručným zopakovaním základných definícií súvisiacich s konečnými automatmi v podobe, v akej odzneli na prednáške. Rovno pritom upozornime na fakt, že ide len o jeden z množstva prístupov k definícii konečných automatov a nimi rozoznávaných jazykov; pri štúdiu literatúry je teda potrebné počítať s určitými rozdielmi v týchto definíciách, ktoré sú však – ako pochopíme neskôr – spravidla relatívne nepodstatné.

Definícia 1. *Deterministický konečný automat* je pätica $A = (K, \Sigma, \delta, q_0, F)$, kde K je konečná množina stavov, Σ je vstupná abeceda, $\delta\colon K \times \Sigma \to K$ je prechodová funkcia, $q_0 \in K$ je počiatočný stav a $F \subseteq K$ je množina koncových (alebo akceptačných) stavov.

Všimnime si, že dôsledkom existencie počiatočného stavu q_0 je neprázdnosť množiny stavov K.

Definícia 2. *Konfigurácia* deterministického konečného automatu $A = (K, \Sigma, \delta, q_0, F)$ je dvojica (q, w), kde $q \in K$ je stav a $w \in \Sigma^*$ je slovo (reprezentujúce nedočítanú časť vstupu).

Definícia 3. *Krok výpočtu* deterministického konečného automatu $A = (K, \Sigma, \delta, q_0, F)$ je binárna relácia $\vdash_A$ na množine konfigurácií automatu A taká, že pre $p, q \in K$ a $u, v \in \Sigma^*$ je $(p, u) \vdash_A (q, v)$ práve vtedy, keď existuje $c \in \Sigma$ také, že $u = cv$ a $\delta(p, c) = q$.

V prípade, že je uvažovaný automat A zrejmý z kontextu, píšeme namiesto $\vdash_A$ často iba $\vdash$. Ide pritom o reláciu; pre každé $k \in \mathbb{N}$ sú teda v relácii $\vdash^k$ tie dvojice konfigurácií, medzi ktorými možno „prejsť“ na k krokov výpočtu. V relácii $\vdash^*$, reflexívno-tranzitívnom uzávere relácie $\vdash$, sú tie dvojice konfigurácií, medzi ktorými možno „prejsť“ na nejaký počet krokov a v tranzitívnom uzávere, relácii $\vdash^+$, sú tie dvojice konfigurácií, medzi ktorými možno „prejsť“ na nejaký nenulový počet krokov. Relácia $\vdash^0$ je identita.

Definícia 4. Nech $A = (K, \Sigma, \delta, q_0, F)$ je deterministický konečný automat. *Jazyk* akceptovaný (alebo rozoznávaný) automatom A je daný ako $L(A) = \{w \in \Sigma^* \mid \exists q \in F : (q_0, w) \vdash^* (q, \varepsilon)\}$.

Poznámka 1. V literatúre sa deterministické konečné automaty niekedy definujú aj s čiastočnou prechodovou funkciou, čo znamená, že hodnota $\delta(q, c)$ nemusí byť definovaná pre všetky stavy q a symboly c.[1] Na tomto predmete však pracujeme s definíciou s úplnou prechodovou funkciou – pri zadávaní deterministického konečného automatu je teda potrebné špecifikovať výstupy jeho prechodovej funkcie pre všetky stavy a všetky písmená.

Nedeterministický konečný automat sa oproti deterministickému líši hlavne tým, že v ňom z jedného stavu môže viesť aj viacero prechodov na ten istý symbol (prípadne nemusí existovať žiaden takýto prechod). To bolo na prednáške sformalizované pomocou prechodovej funkcie, ktorá pre každý stav q a symbol c vráti *množinu* stavov $\delta(q, c)$. Výpočet automatu tak môže zo stavu q na písmeno c pokračovať do všetkých stavov z množiny $\delta(q, c)$, v dôsledku čoho môže na vstupnom slove w existovať viacero rôznych výpočtov. Automat pritom slovo w akceptuje práve vtedy, keď *existuje aspoň jeden* jeho výpočet na slove w, ktorý sa skončí v akceptačnom stave. Pri definícii z prednášky majú navyše nedeterministické konečné automaty možnosť vykonávať aj *prechody na prázdne slovo*, pri ktorých nič neprečítajú zo vstupu.

[1] Ekvivalencia oboch definícií je priamym dôsledkom ekvivalencie deterministických a nedeterministických konečných automatov, ktorej sa budeme venovať nižšie.

Definícia 5. *Nedeterministický konečný automat* je pätica $A = (K, \Sigma, \delta, q_0, F)$, kde K je konečná množina stavov, Σ je vstupná abeceda, $\delta\colon K \times (\Sigma \cup \{\varepsilon\}) \to 2^K$ je prechodová funkcia, $q_0 \in K$ je počiatočný stav a $F \subseteq K$ je množina koncových (alebo akceptačných) stavov.

Definícia 6. *Konfigurácia* nedeterministického konečného automatu $A = (K, \Sigma, \delta, q_0, F)$ je dvojica (q, w), kde $q \in K$ je stav a $w \in \Sigma^*$ je slovo (reprezentujúce nedočítanú časť vstupu).

Definícia 7. *Krok výpočtu* nedeterministického konečného automatu $A = (K, \Sigma, \delta, q_0, F)$ je binárna relácia $\vdash_A$ na množine konfigurácií automatu A taká, že pre $p, q \in K$ a $u, v \in \Sigma^*$ je $(p, u) \vdash_A (q, v)$ práve vtedy, keď existuje $z \in \Sigma \cup \{\varepsilon\}$ také, že $u = zv$ a $q \in \delta(p, z)$.

Podobne ako pri deterministických konečných automatoch píšeme v prípadoch, keď je uvažovaný automat A zrejmý z kontextu, namiesto $\vdash_A$ často iba $\vdash$.

Definícia 8. Nech $A = (K, \Sigma, \delta, q_0, F)$ je nedeterministický konečný automat. *Jazyk* akceptovaný (alebo rozoznávaný) automatom A je daný ako $L(A) = \{w \in \Sigma^* \mid \exists q \in F : (q_0, w) \vdash^* (q, \varepsilon)\}$.

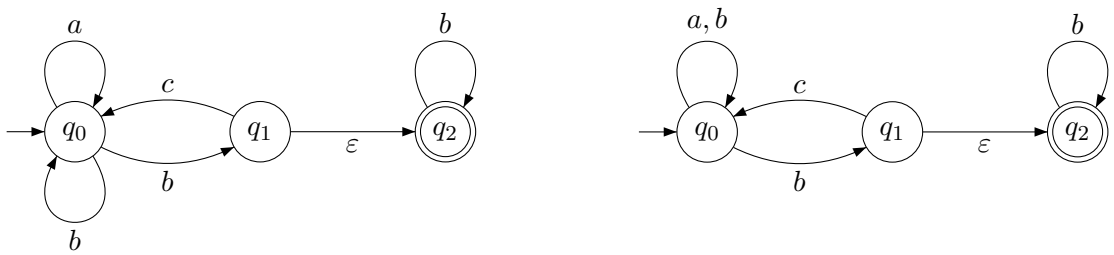
Častým spôsobom ako zadať konečný automat je tzv. *prechodový diagram*, v ktorom sú stavy znázornené „kolečkami“, prechodová funkcia šípkami medzi stavmi, počiatočný stav krátkou šípkou do zodpovedajúceho „kolečka“ a akceptačné stavy „zdvojenými kolečkami“.

Príklad 1. Uvažujme nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ s $K = \{q_0, q_1, q_2\}$, $\Sigma = \{a, b, c\}$, $F = \{q_2\}$ a prechodovou funkciou danou nasledovne:

$$\begin{array}{llll}
\delta(q_0, a) = \{q_0\}, & \delta(q_0, b) = \{q_0, q_1\}, & \delta(q_0, c) = \emptyset, & \delta(q_0, \varepsilon) = \emptyset, \\
\delta(q_1, a) = \emptyset, & \delta(q_1, b) = \emptyset, & \delta(q_1, c) = \{q_0\}, & \delta(q_1, \varepsilon) = \{q_2\}, \\
\delta(q_2, a) = \emptyset, & \delta(q_2, b) = \{q_2\}, & \delta(q_2, c) = \emptyset & \delta(q_2, \varepsilon) = \emptyset.
\end{array}$$

Ľahko vidieť – a pri troche úsilia aj dokázať – že tento automat rozoznáva jazyk $L(A) = \{a, b, bc\}^* b^+$.

Prechodový diagram automatu A je na obrázku 1a. Násobné prechody medzi rovnakými stavmi niekedy pre prehľadnosť kreslíme ako jedinú šípku s viacerými ohodnoteniami, tak ako v prípade slučky v diagrame na obrázku 1b.



(a) Diagram automatu $\mathcal{A}$.

(b) Ekvivalentný diagram toho istého automatu.

Obr. 1: Prechodový diagram nedeterministického konečného automatu A.

Nasledujúce jednoduché tvrdenie, ktorého dôkaz prenechávame ako jednu z úloh na nasledujúce cvičenie, budeme často používať bez toho, aby sme to explicitne uvádzali (platí pre deterministické aj pre nedeterministické konečné automaty):

Tvrdenie 1. *Nech $A = (K, \Sigma, \delta, q_0, F)$ je konečný automat. Potom:*

a) Pre všetky $p, q \in K$ a $u, v \in \Sigma^*$ je $(p, uv) \vdash^* (q, v)$ práve vtedy, keď $(p, u) \vdash^* (q, \varepsilon)$,

b) Pre všetky $p, q, r \in K$ a $u, v \in \Sigma^*$ také, že $(p, u) \vdash^* (q, \varepsilon)$ a $(q, v) \vdash^* (r, \varepsilon)$ je aj $(p, uv) \vdash^* (r, \varepsilon)$.

2 Dokazovanie správnosti konštrukcie automatu

Dokazovanie tvrdenia $L(A) = L$ pre konečný automat A a jazyk L je v mnohom podobné dokazovaniu takýchto tvrdení pre bezkontextové gramatiky. Hlavným spoločným znakom je matematická indukcia, ktorou sa obyčajne dokáže o niečo silnejšie tvrdenie, než to pôvodne zamýšľané. Pre konečné automaty sú takýmto vhodným silnejším tvrdením väčšinou *invarianty* pre jednotlivé stavy: pre každý stav q sa charakterizujú slová w dočítané[2] automatom v stave q – čiže slová w, pre ktoré je $(q_0, w) \vdash^* (q, \varepsilon)$. Tvrdenie $L(A) = L$ potom obyčajne vyplýva z invariantov pre akceptačné stavy.

2.1 Riešené úlohy

Úloha 1. Zostrojte deterministický konečný automat akceptujúci jazyk

$$L = \{w \in \{a, b\}^* \mid \#_a(w) \equiv 1 \pmod 3 \ \wedge \ \#_b(w) \equiv 3 \pmod 7\}$$

a dokážte správnosť svojej konštrukcie.

Riešenie. Dokážeme, že jazyk L je akceptovaný automatom $A = (K, \Sigma, \delta, q_0, F)$, kde $K = \mathbb{Z}_3 \times \mathbb{Z}_7$, $\Sigma = \{a, b\}$, $q_0 = [0, 0]$, $F = \{[1, 3]\}$ a

$$\begin{aligned} \delta([s, t], a) &= [s + 1, t], \quad \forall s \in \mathbb{Z}_3 \ \forall t \in \mathbb{Z}_7, \\ \delta([s, t], b) &= [s, t + 1], \quad \forall s \in \mathbb{Z}_3 \ \forall t \in \mathbb{Z}_7, \end{aligned}$$

kde operácie sčítania sú – keďže ich aplikujeme na prvky grupy $\mathbb{Z}_3$ resp. $\mathbb{Z}_7$ – vždy modulo 3 resp. 7.

Rovnosť $L(A) = L$ vyplynie z nasledujúceho tvrdenia o *invariantoch* pre jednotlivé stavy:

$$\forall [s, t] \in K \ \forall w \in \Sigma^* : ([0, 0], w) \vdash^* ([s, t], \varepsilon) \iff \#_a(w) \equiv s \pmod 3 \ \wedge \ \#_b(w) \equiv t \pmod 7.$$

Dokážme teraz jednotlivé implikácie, vždy naraz pre všetky stavy $[s, t]$ a slová w.

$\Rightarrow$: Indukciou vzhľadom na $n \in \mathbb{N}$ dokážeme, že pre všetky stavy $[s, t] \in K$ a všetky $w \in \Sigma^*$ spĺňajúce $([0, 0], w) \vdash^n ([s, t], \varepsilon)$ musí byť $\#_a(w) \equiv s \pmod 3$ a zároveň $\#_b(w) \equiv t \pmod 7$.

1. Nech $n = 0$. Potom $[s, t] = [0, 0]$ a $w = \varepsilon$. Tvrdenie teda platí.
2. Predpokladajme, že tvrdenie platí pre $n = k$ a uvažujme $n = k + 1$.
   Nech $[s, t] \in K$ a $w \in \Sigma^*$ sú také, že $([0, 0], w) \vdash^{k+1} ([s, t], \varepsilon)$. Potom $w = uc$ pre nejaké $u \in \Sigma^*$ a $c \in \Sigma$ také, že pre nejaký stav $[s', t'] \in K$ je

   $$([0, 0], uc) \vdash^k ([s', t'], c) \vdash ([s, t], \varepsilon)$$

   (posledný krok výpočtu musí byť na písmeno, pretože automat A je deterministický).
   a) Ak $c = a$, nutne $\delta([s', t'], a) = [s, t]$. Z definície prechodovej funkcie δ vyplýva, že $s' = s - 1$ a $t' = t$ a vďaka indukčnému predpokladu dostávame $\#_a(u) \equiv s - 1 \pmod 3$ a $\#_b(u) \equiv t \pmod 7$. Preto naozaj $\#_a(w) = \#_a(ua) = \#_a(u) + 1 \equiv s \pmod 3$ a $\#_b(w) = \#_b(ua) = \#_b(u) \equiv t \pmod 7$.
   b) Ak $c = b$, možno argumentovať analogicky.

$\Leftarrow$: Keďže je automat A deterministický, je možné dokázať opačnú implikáciu aj bez ďalšej indukcie. Stačí si všimnúť, že dokazované invarianty pre jednotlivé stavy sú „disjunktné“ – nemôže sa teda stať, že by niektoré slovo súčasne vyhovovalo invariantom pre dva rôzne stavy – a každé slovo nad abecedou Σ súčasne vyhovuje invariantu pre niektorý zo stavov.

Z úplnosti prechodovej funkcie deterministického konečného automatu A vyplýva, že pre každé slovo $w \in \Sigma^*$ existuje stav $q \in K$ taký, že $([0, 0], w) \vdash^* (q, \varepsilon)$. Nech slovo w vyhovuje invariantu pre nejaký stav $[s, t] \in K$, t. j. $\#_a(w) \equiv s \pmod 3$ a $\#_b(w) \equiv t \pmod 7$. Potom nutne $q = [s, t]$, pretože v opačnom prípade by podľa predchádzajúcej implikácie muselo slovo w vyhovovať invariantu pre stav $q \neq [s, t]$, čo sa vylučuje s platnosťou invariantu pre stav $[s, t]$.

[2] Podobne by bolo možné pre každý stav q charakterizovať aj tie slová, pre ktoré existuje výpočet začínajúci v stave q a končiaci v akceptačnom stave. Takýto prístup je však väčšinou o niečo menej intuitívny.

Rovnosť $L(A) = L$ je už bezprostredným dôsledkom dokázaného:

$\subseteq$: Nech $w \in L(A)$. Keďže $[1,3]$ je jediný akceptačný stav, musí platiť $([0,0], w) \vdash^* ([1,3], \varepsilon)$. Z invariantu pre $[1,3]$ ale vyplýva $\#_a(w) \equiv 1 \pmod 3$ a $\#_b(w) \equiv 3 \pmod 7$, a teda $w \in L$.

$\supseteq$: Nech $w \in L$. Potom $\#_a(w) \equiv 1 \pmod 3$ a $\#_b(w) \equiv 3 \pmod 7$ a z dokázaných invariantov vyplýva, že $([0,0], w) \vdash^* ([1,3], \varepsilon)$. Preto $w \in L(A)$. $\square$

Úloha 2. Nech $\Sigma = \{a, b\}$. Zostrojte nedeterministický konečný automat akceptujúci jazyk

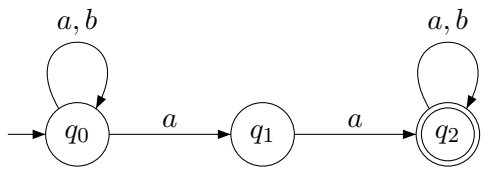
$$L = \Sigma^* aa \Sigma^*$$

a dokážte správnosť svojej konštrukcie.

Riešenie. Ukážeme, že jazyk L je akceptovaný automatom $A = (K, \Sigma, \delta, q_0, F)$, kde $K = \{q_0, q_1, q_2\}$, $F = \{q_2\}$ a

$$\begin{aligned}
&\delta(q_0, a) = \{q_0, q_1\}, && \delta(q_1, a) = \{q_2\}, && \delta(q_2, a) = \{q_2\},\\
&\delta(q_0, b) = \{q_0\}, && \delta(q_1, b) = \emptyset, && \delta(q_2, b) = \{q_2\},\\
&\delta(q_0, \varepsilon) = \emptyset, && \delta(q_1, \varepsilon) = \emptyset, && \delta(q_2, \varepsilon) = \emptyset.
\end{aligned}$$

Prechodový diagram automatu A je na obrázku 2.



Obr. 2: Prechodový diagram nedeterministického konečného automatu A.

Pre stavy automatu A dokážeme nasledujúce invarianty:

$$\begin{aligned}
I_0 &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^* (q_0, \varepsilon) \iff w \in \Sigma^*,\\
I_1 &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^* (q_1, \varepsilon) \iff w \in \Sigma^* a,\\
I_2 &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^* (q_2, \varepsilon) \iff w \in \Sigma^* aa \Sigma^*.
\end{aligned}$$

$\Rightarrow$: Matematickou indukciou vzhľadom na n dokážeme, že pre všetky $n \in \mathbb{N}$:

$$\begin{aligned}
I_0' &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^n (q_0, \varepsilon) \Rightarrow w \in \Sigma^*,\\
I_1' &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^n (q_1, \varepsilon) \Rightarrow w \in \Sigma^* a,\\
I_2' &\colon \forall w \in \Sigma^* : (q_0, w) \vdash^n (q_2, \varepsilon) \Rightarrow w \in \Sigma^* aa \Sigma^*.
\end{aligned}$$

1. Pre $n = 0$ môže byť ľavá strana pravdivá iba pri implikácii I_0'. V takom prípade tiež nutne $w = \varepsilon$, čo je slovo zo Σ^*.
2. Predpokladajme platnosť implikácií I_0' až I_2' pre $n = k$ a uvažujme $n = k + 1$.

   I_0': Pre každé $w \in \Sigma^*$ spĺňajúce $(q_0, w) \vdash^{k+1} (q_0, \varepsilon)$ je triviálne $w \in \Sigma^*$.

   I_1': Nech $w \in \Sigma^*$ je slovo také, že $(q_0, w) \vdash^{k+1} (q_1, \varepsilon)$. Potom $(q_0, uz) \vdash^k (q, z) \vdash (q_1, \varepsilon)$ pre nejaké $q \in K$, $u \in \Sigma^*$ a $z \in \Sigma \cup \{\varepsilon\}$ také, že $w = uz$.
   V takom prípade $q_1 \in \delta(q, z)$ a z definície prechodovej funkcie δ ľahko vidieť, že $z = a$. Preto naozaj $w = ua \in \Sigma^* a$.

   I_2': Nech $w \in \Sigma^*$ je slovo také, že $(q_0, w) \vdash^{k+1} (q_2, \varepsilon)$. Potom $(q_0, uz) \vdash^k (q, z) \vdash (q_2, \varepsilon)$ pre nejaké $q \in K$, $u \in \Sigma^*$ a $z \in \Sigma \cup \{\varepsilon\}$ také, že $w = uz$.
   Preto $q_2 \in \delta(q, z)$, a teda buď $q = q_1$ a $z = a$, alebo $q = q_2$ a $z \in \{a, b\}$. V prvom prípade z indukčného predpokladu dostávame $u \in \Sigma^* a$, z čoho $w = ua \in \Sigma^* aa \Sigma^*$. V druhom je z indukčného predpokladu $u \in \Sigma^* aa \Sigma^*$, a teda aj $w = uz \in \Sigma^* aa \Sigma^*$.

$\Leftarrow$: Nech $w \in \Sigma^*$. Dokážme najprv implikáciu

$$I_0'' \colon w \in \Sigma^* \Rightarrow (q_0, w) \vdash^* (q_0, \varepsilon).$$

Indukciou vzhľadom na dĺžku slova w.

1. Ak $|w| = 0$, je $w = \varepsilon$ a $(q_0, \varepsilon) \vdash^* (q_0, \varepsilon)$.
2. Predpokladajme platnosť implikácie pre slová dĺžky k a uvažujme $w \in \Sigma^*$ s $|w| = k + 1$. Nech $w = uc$, kde $u \in \Sigma^k$ a $c \in \Sigma$. Na slovo u sa vzťahuje indukčný predpoklad, a teda $(q_0, u) \vdash^* (q_0, \varepsilon)$. Keďže ale $q_0 \in \delta(q_0, a)$ a $q_0 \in \delta(q_0, b)$, nutne $(q_0, c) \vdash (q_0, \varepsilon)$. V dôsledku toho skutočne $(q_0, w) = (q_0, uc) \vdash^* (q_0, c) \vdash (q_0, \varepsilon)$.

Dokážeme teraz druhú implikáciu

$$I_1'' \colon w \in \Sigma^* a \Rightarrow (q_0, w) \vdash^* (q_1, \varepsilon).$$

Každé slovo $w \in \Sigma^* a$ môžeme napísať ako $w = ua$, kde $u \in \Sigma^*$. Podľa implikácie I_0'' teda $(q_0, u) \vdash^* (q_0, \varepsilon)$. Keďže ale $q_1 \in \delta(q_0, a)$, dostávame $(q_0, w) = (q_0, ua) \vdash^* (q_0, a) \vdash (q_1, \varepsilon)$.

Zostáva dokázať poslednú implikáciu

$$I_2'' \colon w \in \Sigma^* aa \Sigma^* \Rightarrow (q_0, w) \vdash^* (q_2, \varepsilon).$$

Slovo $w \in \Sigma^* aa \Sigma^*$ napíšme ako $w = uaav$ pre nejaké slová $u, v \in \Sigma^*$. Implikáciu I_2'' dokážeme indukciou vzhľadom na dĺžku slova v.

1. Pre $|v| = 0$ je $v = \varepsilon$ a $w = uaa$. Podľa implikácie I_1'' je $(q_0, ua) \vdash^* (q_1, \varepsilon)$. Keďže ale $q_2 \in \delta(q_1, a)$, dostávame $(q_0, w) = (q_0, uaa) \vdash^* (q_1, a) \vdash (q_2, \varepsilon)$.
2. Predpokladajme platnosť tvrdenia pre $|v| = k$ a uvažujme $v \in \Sigma^*$ také, že $|v| = k + 1$. Potom $w = uaaxc$, kde $x \in \Sigma^k$ a $c \in \Sigma$. Z indukčného predpokladu $(q_0, uaax) \vdash^* (q_2, \varepsilon)$. Keďže ale $q_2 \in \delta(q_2, a)$ a $q_2 \in \delta(q_2, b)$, dostávame $(q_0, w) = (q_0, uaaxc) \vdash^* (q_2, c) \vdash (q_2, \varepsilon)$.

Keďže je q_2 jediným akceptačným stavom, $w \in \Sigma^*$ patrí do $L(A)$ práve vtedy, keď $(q_0, w) \vdash^* (q_2, \varepsilon)$. To ale vďaka invariantu I_2 nastane práve vtedy, keď $w \in \Sigma^* aa \Sigma^* = L$. Skutočne teda $L(A) = L$. $\square$

3 „Odepsilonovanie“ nedeterministických konečných automatov

Na prednáške bolo dokázané, že prechody na prázdne slovo nie sú pre nedeterministické konečné automaty nevyhnutné – každý nedeterministický konečný automat možno prerobiť na ekvivalentný, ktorý žiadne takéto prechody nemá.

Veta 1. *Nech $A = (K, \Sigma, \delta, q_0, F)$ je nedeterministický konečný automat. Potom existuje nedeterministický konečný automat $A' = (K', \Sigma', \delta', q_0', F')$ taký, že $L(A') = L(A)$ a pre všetky $q \in K'$ je $\delta'(q, \varepsilon) = \emptyset$.*

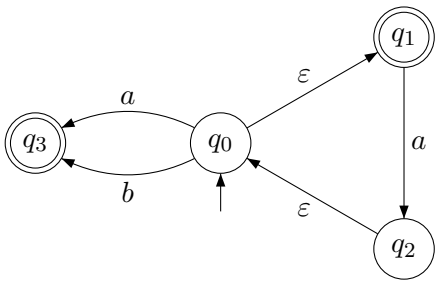
Konštrukcií zbavujúcich nedeterministický konečný automat prechodov na prázdne slovo existuje viacero. V nasledujúcom si pripomenieme algoritmus, ktorého správnosť bola dokázaná na prednáške. Jeho vstupom je nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$.

1. Pre všetky $q \in K$ nájdi (napríklad prehľadávaním do hĺbky) „epsilonový chvost stavu q“, čo je množina $[q]_\varepsilon$ stavov dosiahnuteľných z q na prázdne slovo: $[q]_\varepsilon = \{p \in K \mid (q, \varepsilon) \vdash^* (p, \varepsilon)\}$.
2. Odstráň všetky prechody na ε.
3. Pre každý stav $q \in K$ okrem q_0 a každé $c \in \Sigma$ pridaj prechody zo stavu q na písmeno c do všetkých stavov $p \in K$ takých, že pre nejaké $q' \in K$ je $q' \in \delta(q, c)$ a $p \in [q']_\varepsilon$.
4. Pre každé $c \in \Sigma$ pridaj prechody zo stavu q_0 na písmeno c do všetkých stavov $p \in K$ takých, že pre nejaké $q, q' \in K$ je $q \in [q_0]_\varepsilon$, $q' \in \delta(q, c)$ a $p \in [q']_\varepsilon$.
5. Ak $[q_0]_\varepsilon$ obsahuje aspoň jeden akceptačný stav, pridaj q_0 do množiny akceptačných stavov.

Poznámka 2. V krokoch 3 a 4 uvedeného algoritmu označuje δ prechodovú funkciu *pôvodného* automatu A. Napríklad v kroku 4 teda netreba uvažovať prechody pridané v kroku 3.

3.1 Riešená úloha

Úloha 3. Uvažujme nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ s $K = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{a, b\}$, $F = \{q_1, q_3\}$ a prechodovou funkciou δ danou prechodovým diagramom na obrázku 3. Štandardnou konštrukciou zbavte automat A prechodov na ε.

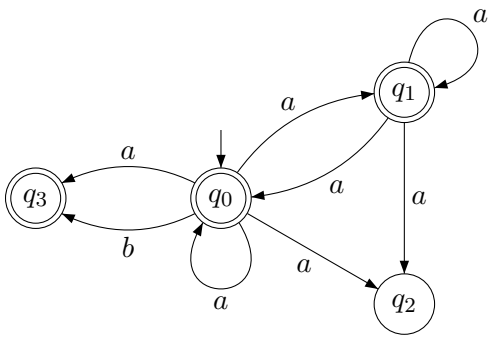


Obr. 3: Prechodový diagram nedeterministického konečného automatu A.

Riešenie. Pre každý stav $q \in K$ najprv nájdeme jeho „epsilonový chvost“ $[q]_\varepsilon$:

$$[q_0]_\varepsilon = \{q_0, q_1\}, \qquad [q_1]_\varepsilon = \{q_1\},$$
$$[q_2]_\varepsilon = \{q_2, q_0, q_1\}, \qquad [q_3]_\varepsilon = \{q_3\}.$$

V kroku 2 odstránime prechody na ε vedúce z q_0 do q_1 a z q_2 do q_0. Pokračujme krokom 3. Keďže $q_2 \in \delta(q_1, a)$ a $q_0, q_1 \in [q_2]_\varepsilon$, vo výslednom automate budú prechody zo stavu q_1 na písmeno a vedúce do stavov q_0 a q_1. V kroku 3 nepribudnú žiadne ďalšie prechody. Keďže $q_1 \in [q_0]_\varepsilon$, $q_2 \in \delta(q_1, a)$ a $q_0, q_1, q_2 \in [q_2]_\varepsilon$, v kroku 4 pribudnú prechody zo stavu q_0 na písmeno a do stavov q_0, q_1 a q_2.



Obr. 4: Prechodový diagram nedeterministického konečného automatu A'.

V poslednom kroku sa stane akceptačným aj stav q_0, pretože $q_1 \in [q_0]_\varepsilon \cap F$. Výsledný automat $A' = (K', \Sigma', \delta', q_0', F')$ je potom znázornený diagramom na obrázku 4. □

Poznámka 3. Algoritmus z prednášky je v zásade založený na nahradzovaní postupností prechodov typu „jeden prechod na písmeno a niekoľko prechodov na ε“ jedným prechodom na písmeno. Existuje aj duálny algoritmus, v ktorom sa jedným prechodom na písmeno nahradzujú postupnosti typu „niekoľko prechodov na ε a jeden prechod na písmeno“. Takýto prístup umožňuje spracovať počiatočný stav q_0 konzistentne s ostatnými stavmi – nie je teda nutná analógia kroku 4. Krok 5 by ale naopak bolo nutné vykonať pre všetky stavy (a nielen pre počiatočný stav q_0).

Obidva tieto varianty „odepsilonovacieho“ algoritmu možno upraviť aj pre nedeterministické konečné automaty s viac ako jedným počiatočným stavom, kde ich možno interpretovať azda najprirodzenejším a najelegantnejším spôsobom.

4 Determinizácia konečných automatov

Je jedným z kľúčových výsledkov teórie konečných automatov, že deterministické a nedeterministické konečné automaty sú rovnako silné. Jeden smer tohto tvrdenia je zrejmý: aj keď deterministický konečný automat v zmysle jeho formálnej definície z prednášky *nie je* súčasne aj nedeterministickým konečným automatom, po drobnej úprave prechodovej funkcie sa ním stane. Zvyšná časť tvrdenia bola dokázaná na prednáške a možno ju sformulovať v podobe nasledujúcej vety.

Veta 2. *Nech A je nedeterministický konečný automat. Potom existuje* deterministický *konečný automat A' taký, že $L(A') = L(A)$.*

4.1 Podmnožinová konštrukcia

Algoritmus determinizácie nedeterministického konečného automatu predpokladá ako svoj vstup už „odepsilonovaný“ nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ a spočíva v takzvanej podmnožinovej konštrukcii:

1. Nech $K' = 2^K$.
2. Pre každé $Q \in K'$ a každé $c \in \Sigma$ polož $\delta'(Q, c) = \{p \in K \mid \exists q \in Q : p \in \delta(q, c)\}$.
3. Za počiatočný stav q'_0 vezmi množinu $\{q_0\}$.
4. Za F' vezmi množinu všetkých stavov $Q \in K'$ takých, že $Q \cap F \neq \emptyset$.

Výstupom je deterministický konečný automat $A' = (K', \Sigma, \delta', q'_0, F')$ taký, že $L(A') = L(A)$.

4.2 Zefektívnenie podmnožinovej konštrukcie

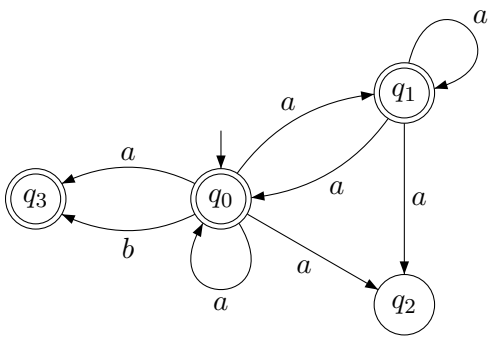
Výsledkom podmnožinovej konštrukcie v podobe, v akej bola opísaná vyššie, môže vo všeobecnosti byť aj automat s veľmi veľkým množstvom nedosiahnuteľných – a teda zbytočných – stavov. Tomuto nedostatku možno predísť *postupným konštruovaním* ekvivalentného deterministického automatu A' tak, aby tento obsahoval *iba dosiahnuteľné* stavy. Vstupom nasledujúcej schémy determinizačného algoritmu[3] je opäť už „odepsilonovaný“ nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ a jej výstupom bude deterministický konečný automat $A' = (K', \Sigma, \delta', q'_0, F')$ taký, že $L(A') = L(A)$. V priebehu vykonávania bude množina $K'_?$ pozostávať z tých stavov konštruovaného automatu, ktoré už boli pridané do množiny K', ale ešte pre ne nie sú definované výstupy prechodovej funkcie δ'.

1. Inicializuj $K' := \{\{q_0\}\}$ a $K'_? := \{\{q_0\}\}$.
2. Vyber ľubovoľné $Q \in K'_?$ a polož $K'_? := K'_? - \{Q\}$.
3. Opakuj pre všetky $c \in \Sigma$:
   3.1. Nech $P := \{p \in K \mid \exists q \in Q : p \in \delta(q, c)\}$.
   3.2. Ak $P \notin K'$, polož $K' := K' \cup \{P\}$ a $K'_? := K'_? \cup \{P\}$.
   3.3. Polož $\delta'(Q, c) := P$.
4. Ak $K'_? \neq \emptyset$, pokračuj krokom 2; inak pokračuj krokom 5.
5. Polož $q'_0 := \{q_0\}$ a $F' := \{Q \in K' \mid Q \cap F \neq \emptyset\}$.

[3] O algoritmus by išlo až po upresnení spôsobu vyberania stavov z množiny $K'_?$.

4.3 Riešená úloha

Úloha 4. Nech $A = (K, \Sigma, \delta, q_0, F)$ so $\Sigma = \{a, b\}$ je nedeterministický konečný automat na obrázku 5. Štandardnou konštrukciou zostrojte ekvivalentný *deterministický* konečný automat.



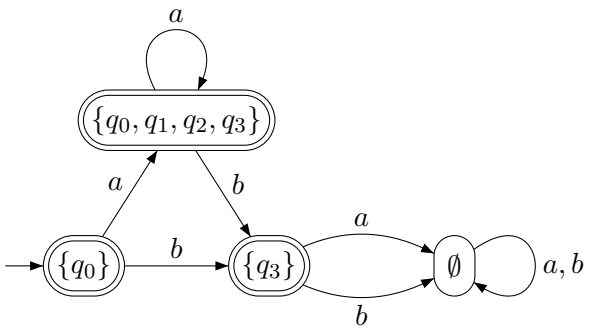
Obr. 5: Prechodový diagram nedeterministického konečného automatu A.

Riešenie. Automat A neobsahuje žiadne prechody na ε. Môžeme naň teda aplikovať podmnožinovú konštrukciu vo variante z pododdielu 4.2 a stavy ekvivalentného automatu A' vytvárať spoločne s jeho prechodmi napríklad nasledovne.

Stav $Q \in K'$	$\delta'(Q, a)$	$\delta'(Q, b)$
$\{q_0\}$	$\{q_0, q_1, q_2, q_3\}$	$\{q_3\}$
$\{q_0, q_1, q_2, q_3\}$	$\{q_0, q_1, q_2, q_3\}$	$\{q_3\}$
$\{q_3\}$	$\emptyset$	$\emptyset$
$\emptyset$	$\emptyset$	$\emptyset$

Počiatočným stavom automatu A' bude množina $\{q_0\}$ a koncovými stavmi budú $\{q_0\}$, $\{q_0, q_1, q_2, q_3\}$ a $\{q_3\}$ – všetky tieto množiny totiž obsahujú aspoň jeden koncový stav automatu A.

Zisťujeme teda, že deterministický konečný automat $A' = (K', \Sigma, \delta', q_0', F')$ ekvivalentný automatu A je daný diagramom na obrázku 6.



Obr. 6: Prechodový diagram deterministického konečného automatu A' ekvivalentného automatu A.

To znamená, že $K' = \{\{q_0\}, \{q_0, q_1, q_2, q_3\}, \{q_3\}, \emptyset\}$,

$$\delta'(\{q_0\}, a) = \{q_0, q_1, q_2, q_3\}, \qquad \delta'(\{q_0\}, b) = \{q_3\},$$
$$\delta'(\{q_0, q_1, q_2, q_3\}, a) = \{q_0, q_1, q_2, q_3\}, \qquad \delta'(\{q_0, q_1, q_2, q_3\}, b) = \{q_3\},$$
$$\delta'(\{q_3\}, a) = \emptyset, \qquad \delta'(\{q_3\}, b) = \emptyset,$$
$$\delta'(\emptyset, a) = \emptyset, \qquad \delta'(\emptyset, b) = \emptyset,$$

$q_0' = \{q_0\}$ a $F' = \{\{q_0\}, \{q_0, q_1, q_2, q_3\}, \{q_3\}\}$. □

5 Konštrukcia regulárnej gramatiky ku konečnému automatu

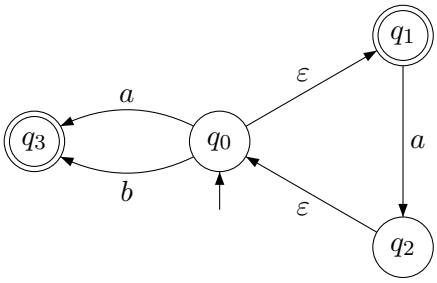
Na prednáške bola dokázaná ekvivalencia (deterministických alebo nedeterministických) konečných automatov a regulárnych gramatík.

Zamerajme sa teraz na konštrukciu regulárnej gramatiky $G = (N, T, P, \sigma)$ ekvivalentnej danému nedeterministickému konečnému automatu $A = (K, \Sigma, \delta, q_0, F)$, ktorú možno zhrnúť nasledovne.

1. Abeceda neterminálov gramatiky G je množina stavov automatu A, t. j. $N = K$.
2. Abeceda terminálov gramatiky G je identická so vstupnou abecedou automatu A, t. j. $T = \Sigma$.
3. Množina prepisovacích pravidiel P gramatiky G je daná takto:
   - (i) Pre každé $p, q \in K$ a $z \in \Sigma \cup \{\varepsilon\}$ také, že $q \in \delta(p, z)$, je v množine P pravidlo $p \to zq$.
   - (ii) Pre každé $q \in F$ je v množine prepisovacích pravidiel P pravidlo $q \to \varepsilon$.
   - (iii) Množina P neobsahuje žiadne iné prepisovacie pravidlá.
4. Počiatočný neterminál gramatiky G je počiatočný stav automatu A, t. j. $\sigma = q_0$.

5.1 Riešená úloha

Úloha 5. Nech $A = (K, \Sigma, \delta, q_0, F)$ so $\Sigma = \{a, b\}$ je nedeterministický konečný automat na obrázku 7.



Obr. 7: Prechodový diagram nedeterministického konečného automatu A.

Štandardnou konštrukciou zostrojte regulárnu gramatiku G ekvivalentnú automatu A.

Riešenie. Položme $G = (N, T, P, \sigma)$, kde $N = \{q_0, q_1, q_2, q_3\}$, $T = \{a, b\}$, $\sigma = q_0$ a

$$\begin{aligned} P = \{ & q_0 \to q_1 \mid aq_3 \mid bq_3 \\ & q_1 \to aq_2 \mid \varepsilon \\ & q_2 \to q_0 \\ & q_3 \to \varepsilon\}. \end{aligned}$$

□

6 Konštrukcia konečného automatu k regulárnej gramatike

Pripomeňme si ešte opačný problém konštrukcie nedeterministického konečného automatu ekvivalentného danej regulárnej gramatike. Konštrukciu automatu $A = (K, \Sigma, \delta, q_0, F)$ ku gramatike $G = (N, T, P, \sigma)$ možno zhrnúť nasledovne.

1. Zostroj regulárnu gramatiku $G' = (N', T, P', \sigma)$ takú, že $P' \subseteq N' \times (TN' \cup N' \cup T \cup \{\varepsilon\})$ a $L(G') = L(G)$. Pre každé pravidlo $\pi = (\alpha \to a_1 \dots a_k s) \in P$, kde $k \geq 2$, $a_1, \dots, a_k \in T$ a $s \in N \cup \{\varepsilon\}$ teda:

   1.1. Zaveď nové neterminály $\psi_{\pi,1}, \dots, \psi_{\pi,k-1}$.

   1.2. Odober pôvodné pravidlo $\alpha \to a_1 \dots a_k s$.

   1.3. Pridaj nové pravidlá $\alpha \to a_1\psi_{\pi,1}, \psi_{\pi,1} \to a_2\psi_{\pi,2}, \dots, \psi_{\pi,k-1} \to a_k s$.

2. Polož $K = N' \cup \{q_{\text{fin}}\}$, kde $q_{\text{fin}} \notin N'$.

3. Polož $\Sigma = T$.

4. Prechodovú funkciu δ automatu A definuj pre každé $\xi \in N'$ a $z \in T \cup \{\varepsilon\}$ nasledovne:

   (i) Ak pre $\eta \in N'$ je $\xi \to z\eta \in P'$, tak $\eta \in \delta(\xi, z)$.

   (ii) Ak $\xi \to z \in P'$, tak $q_{\text{fin}} \in \delta(\xi, z)$.

   (iii) Množina $\delta(\xi, z)$ neobsahuje žiadne iné stavy.

   Navyše $\delta(q_{\text{fin}}, z) = \emptyset$ pre všetky $z \in T \cup \{\varepsilon\}$.

5. Polož $q_0 = \sigma$.

6. Polož $F = \{q_{\text{fin}}\}$.

6.1 Riešená úloha

Úloha 6. Nech $G = (N, T, P, \sigma)$ je regulárna gramatika s $N = \{\sigma, \alpha, \beta, \gamma\}$, $T = \{a, b\}$ a

$$\begin{aligned} P = \{&\sigma \to aba\sigma \mid b\alpha \mid \beta \mid b \\ &\alpha \to a\alpha \mid \varepsilon \\ &\beta \to \beta \mid aa\}. \end{aligned}$$

Štandardnou konštrukciou zostrojte ekvivalentný nedeterministický konečný automat.

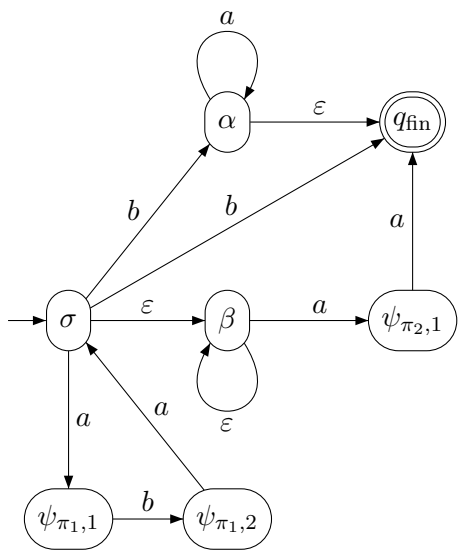
Riešenie. Zaveďme nasledujúce označenia pravidiel gramatiky G:

$$\begin{aligned} \pi_1 &:= (\sigma \to aba\sigma), \\ \pi_2 &:= (\beta \to aa). \end{aligned}$$

Ekvivalentnú regulárnu gramatiku $G' = (N', T, P', \sigma)$ takú, že $P' \subseteq N' \times (TN' \cup N' \cup T \cup \{\varepsilon\})$, zostrojíme nasledovne: $N' = \{\sigma, \alpha, \beta, \psi_{\pi_1,1}, \psi_{\pi_1,2}, \psi_{\pi_2,1}\}$ a

$$\begin{aligned} P' = \{&\sigma \to a\psi_{\pi_1,1} \mid b\alpha \mid \beta \mid b \\ &\alpha \to a\alpha \mid \varepsilon \\ &\beta \to \beta \mid a\psi_{\pi_2,1} \\ &\psi_{\pi_1,1} \to b\psi_{\pi_1,2} \\ &\psi_{\pi_1,2} \to a\sigma \\ &\psi_{\pi_2,1} \to a\}. \end{aligned}$$

Výsledný nedeterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ je potom daný ako na obrázku 8.



Obr. 8: Nedeterministický konečný automat A ekvivalentný gramatike G.

To znamená, že $K = \{\sigma, \alpha, \beta, \psi_{\pi_1,1}, \psi_{\pi_1,2}, \psi_{\pi_2,1}, q_{\text{fin}}\}$, $\Sigma = \{a, b\}$,

$$\begin{aligned}
\delta(\sigma, a) &= \{\psi_{\pi_1,1}\}, & \delta(\sigma, b) &= \{\alpha, q_{\text{fin}}\}, & \delta(\sigma, \varepsilon) &= \{\beta\},\\
\delta(\alpha, a) &= \{\alpha\}, & \delta(\alpha, b) &= \emptyset, & \delta(\alpha, \varepsilon) &= \{q_{\text{fin}}\},\\
\delta(\beta, a) &= \{\psi_{\pi_2,1}\}, & \delta(\beta, b) &= \emptyset, & \delta(\beta, \varepsilon) &= \{\beta\},\\
\delta(\psi_{\pi_1,1}, a) &= \emptyset, & \delta(\psi_{\pi_1,1}, b) &= \{\psi_{\pi_1,2}\}, & \delta(\psi_{\pi_1,1}, \varepsilon) &= \emptyset,\\
\delta(\psi_{\pi_1,2}, a) &= \{\sigma\}, & \delta(\psi_{\pi_1,2}, b) &= \emptyset, & \delta(\psi_{\pi_1,2}, \varepsilon) &= \emptyset,\\
\delta(\psi_{\pi_2,1}, a) &= \{q_{\text{fin}}\}, & \delta(\psi_{\pi_2,1}, b) &= \emptyset, & \delta(\psi_{\pi_2,1}, \varepsilon) &= \emptyset,
\end{aligned}$$

$q_0 = \sigma$ a $F = \{q_{\text{fin}}\}$. $\square$