

Nedeterminizmus v konečných automatoch

študent: Ján Rosina
vedúci: prof. RNDr. Branislav Rován, PhD.

25. júna 2018

- ▶ Skúmanie konečných automatov
 - ▶ Oblasť dlhodobo aktívneho výskumu od 50.-tich rokov
 - ▶ Napriek jednoduchosti konečných automatov pomerne široké spektrum aplikácií
 - ▶ napr. lexikálna analýza
 - ▶ V súčasnosti sa skúma najmä vplyv rôznych operácií na regulárnych jazykoch na stavovú zložitosť k nim konštruovaných konečných automatov

Cieľ práce

- ▶ Skúmanie vzťahu stavovej zložitosti nedeterministického a ekvivalentného deterministického konečného automatu
 - ▶ Vo všeobecnosti je potrebných a postačujúcich 2^n stavov DKA k ekvivalentnému n -stavovému NKA (štandardná konštrukcia)
- ▶ Cieľ práce: hľadanie podtried regulárnych jazykov s menším nárastom stavovej zložitosti DKA oproti NKA

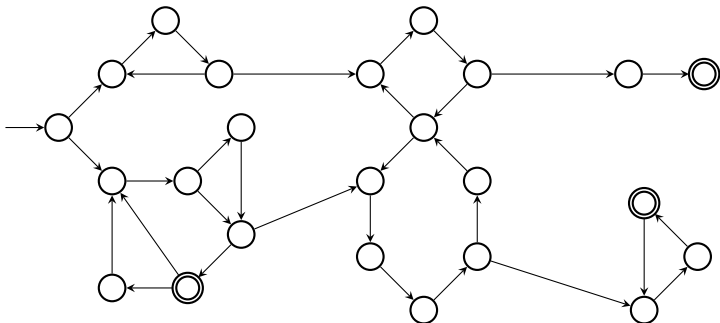
Jazyky definované podslovom

- ▶ Jazyky tvaru $L_x = \{uxv \mid u, v \in \Sigma^*\}$ pre dané slovo x
- ▶ Pre ľubovoľný L_x je stavová zložitosť NKA a DKA akceptujúcich L_x rovnaká
 - ▶ DKA postačí $|x| + 1$ stavov
 - ▶ NKA potrebuje aspoň $|x| + 1$ stavov

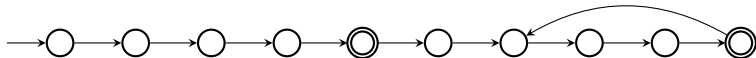
Jazyky nad unárnou abecedou - úvod

- ▶ Jazyky tvaru $L \subseteq \{a\}^*$
- ▶ Predošlý výskum: M. Chrobák dokázal, že vo všeobecnosti je pre daný n -stavový NKA, akceptujúceho jazyk nad unárnou abecedou, na konštrukciu ekvivalentného DKA potrebných a postačujúcich $O(e^{\sqrt{n \log(n)}})$ stavov.

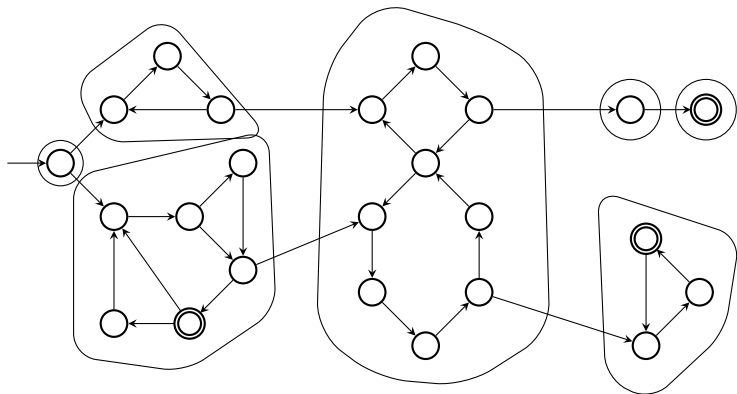
Zatiaľ čo NKA môže obsahovať spleť cyklov,



DKA môže obsahovať len jediný cyklus

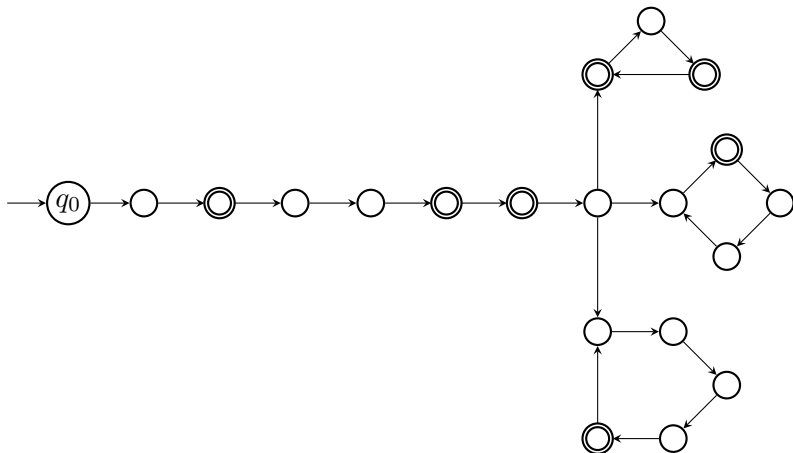


Stavy NKA môžeme rozdeliť na niekoľko silne súvislých komponentov



Normálny tvar definovaný Chrobákom

- Prejdené cykly $C_1, \dots, C_k$ môžeme nahradiť cyklom dĺžky $\gcd(|C_1|, \dots, |C_k|)$
- Ku každému n -stavovému NKA existuje NKA v normálnom tvare s rádovo n^2 stavmi



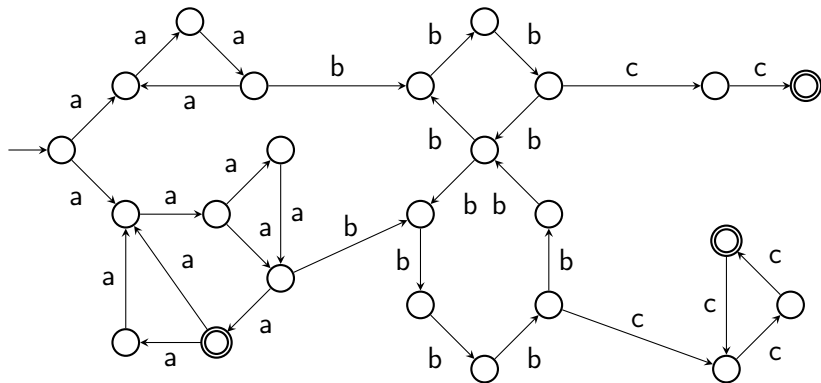
Jazyky nad unárnou abecedou - výsledky

- ▶ Počet stavov DKA závisí od rozvetvenosti grafu silne súvislých komponentov NKA
 - ▶ Ak tento graf obsahuje jedinú vetvu, potom vieme zostrojiť DKA s $O(n^2)$ stavmi
 - ▶ Pre málo rozvetvený graf dostávame prakticky polynomiálnu stavovú zložitosť DKA oproti NKA

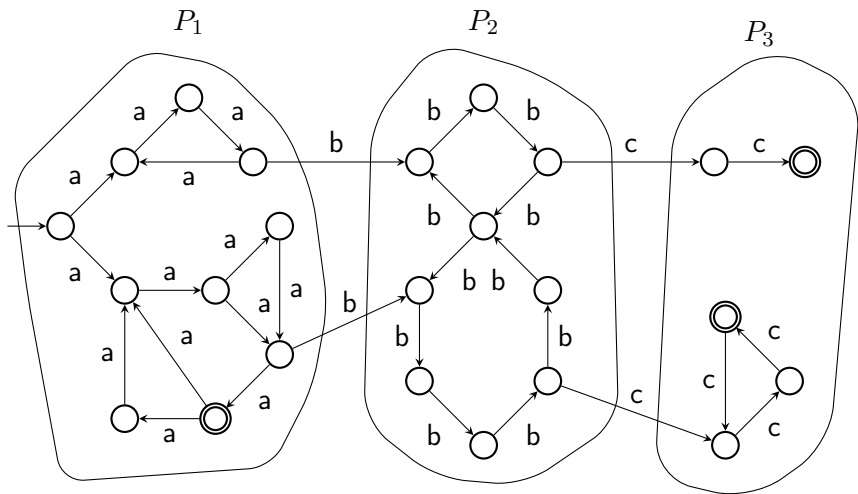
Ohraničené jazyky

- ▶ L je ohraničený jazyk, ak existujú slová $w_1, \dots, w_n$ také, že $L \subseteq w_1^* \dots w_n^*$
- ▶ BÚNV $|w_1| = \dots = |w_n| = 1$
 - ▶ dôsledok kapitoly o jazykoch definovaných podslovom

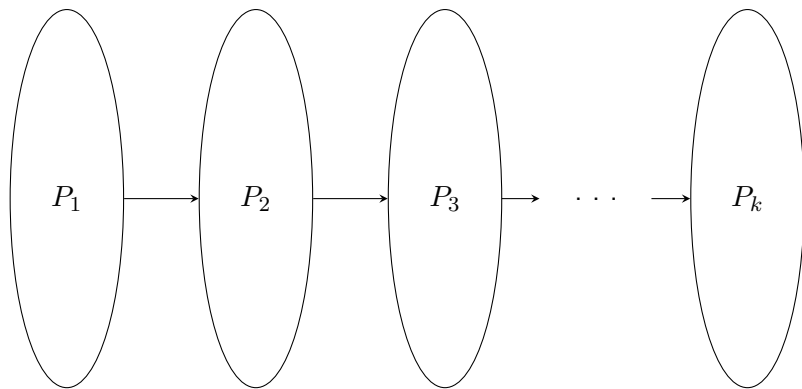
NKA akceptující ohraničený jazyk



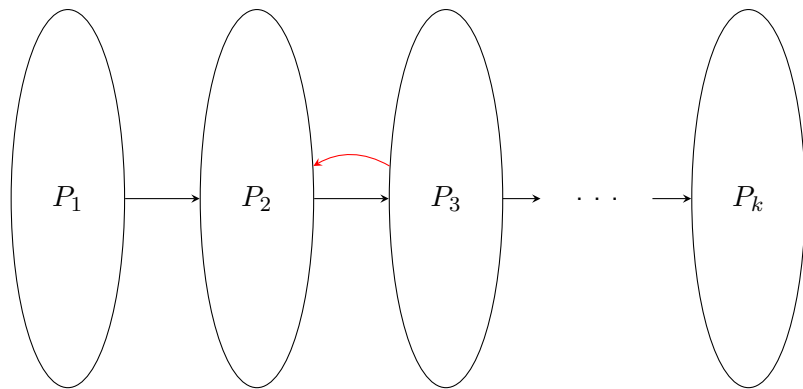
Stavy môžeme rozdeliť na partície



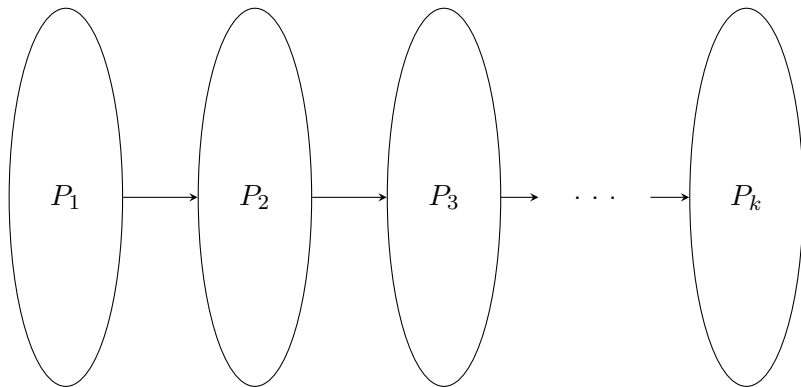
Automat prechádza partíciami jedným smerom...



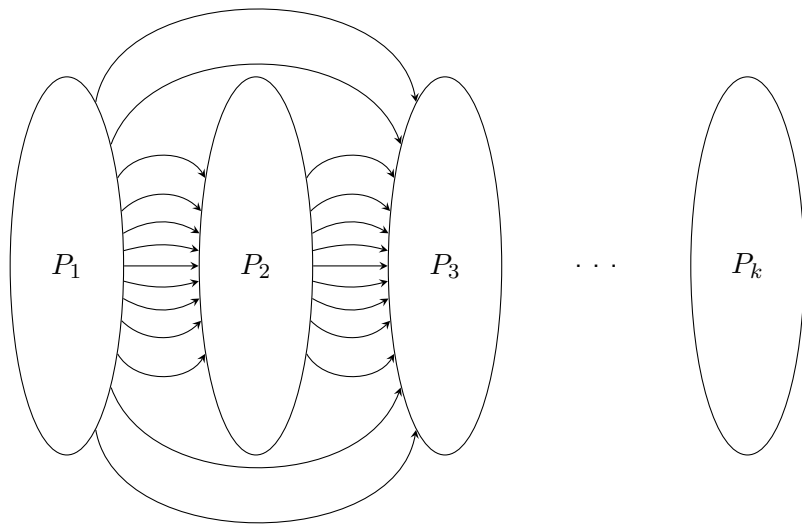
...a nevracia sa do prejdenných partícií



Aká bude celková stavová zložitosť DKA?
Súčet stavov simulovaných partícií?

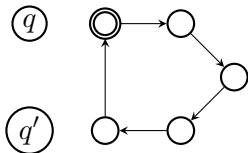


V skutočnosti môžu byť prechody medzi partíciami komplikovanejšie...

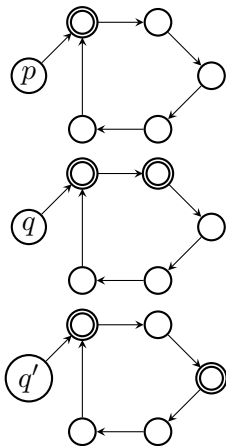
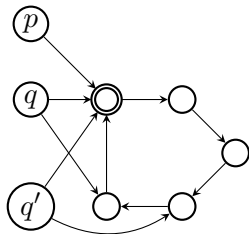


Idea príkladu

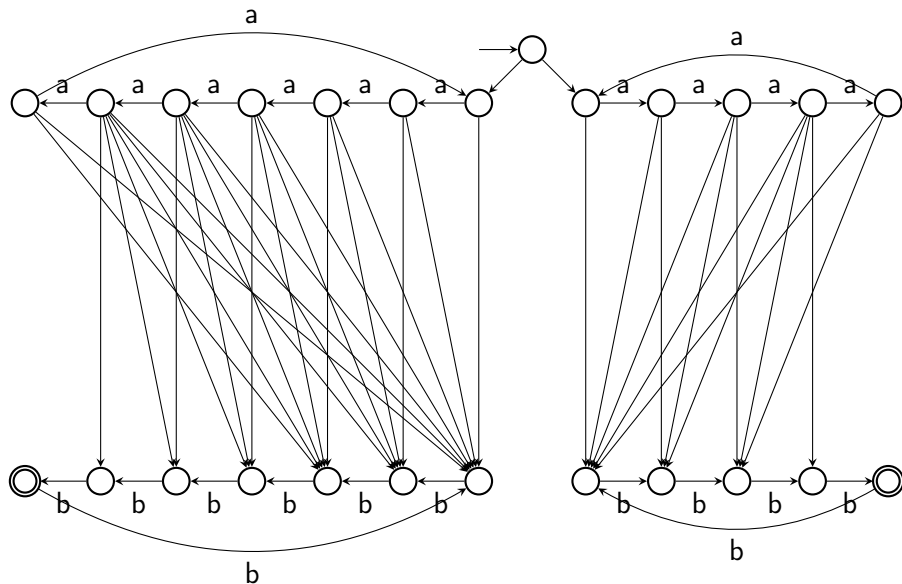
ako spôsobiť potrebu simulovať cyklus viac krát?



Idea príkladu



Konstrukcia príkladu pre cykly dĺžky 5 a 7



Ohraničené jazyky

- ▶ Vo všeobecnosti pre n -stavový NKA s k partíciami postačuje $ke^{k\sqrt{n\log(n)}}$ stavov na zostrojenie ekvivalentného DKA
 - ▶ Zdá sa to veľa, ale k je výrazne menšie ako n
- ▶ Ak obmedzíme prechody medzi partíciami NKA tak, že medzi každými dvoma povolíme najviac jeden, potom na zostrojenie ekvivalentného DKA postačí $\frac{k(k+1)e^{\sqrt{n\log(n)}}}{2}$ stavov

Ďakujem za pozornosť

- ▶ Lema 2.3: Nech je dané ľubovoľné slovo x , nech $A_1 = (K_1, \Sigma_1, \delta_1, q_{0,1}, F_1)$ je nedeterministický konečný automat taký, že $L(A_1) = L_x$, potom existuje deterministický konečný automat $A_2 = (K_2, \Sigma_2, \delta_2, q_{0,2}, F_2)$ taký, že $L(A_2) = L_x$ a súčasne $|K_2| = |K_1|$.
 - ▶ Oprava: Nech ... a súčasne $|K_2| \leq |K_1|$.
- ▶ Dôkaz Vety 3.2: vyjadrovanie dĺžky výpočtu pomocou prejdených cyklov: $|M| = x_0 + x_1|C_1| + \dots + x_r|C_r|$, kde M je postupnosť stavov vo výpočte (tie ho jednoznačne určujú), x_i je počet výskytov ukončeného cyklu C_i a x_0 je počet stavov ktoré ostanú v M po vynechaní ukončených cyklov.
 - ▶ Vyjadrujeme konkrétny výpočet, preto stav, v ktorom výpočet skončí je jednoznačne daný, dva cykly rovnakej dĺžky $|C_k| = |C_l|$ (BÚNV $k < l$) preto môžeme pokladať za rovnaké, teda môžeme zápis zjednodušiť: $|M| = x_0 + x_1|C_1| + \dots + (x_k + x_l)|C_k| + \dots + x_{l-1}|C_{l-1}| + x_{l+1}|C_{l+1}| + \dots + x_r|C_r|$. Takýmto zjednodušením pre všetky dvojice cyklov rovnakej dĺžky dostaneme vyjadrenie dĺžky výpočtu pomocou po dvoch rozdielnych cyklov, teda môžeme aplikovať Dôsledok 3.1.1

- Dôkaz Vety 3.2: Automat A''' konštruujeme tak, že každý cyklus má samostatnú reťaz.

