Prezentácia z predmetu
Formálne Jazyky a Automaty

Tomáš Petrášek

February 2, 2001

Zadanie:

Zostrojte zásobníkový automat A pre jazyk:
L = $\{ucv | u, v \in 1\{0,1\}^*$ a u, v nie sú binárne zápisy dvoch po sebe idúcich čísel takých, že $u \leq v$ a c je terminálny symbol$\}$

Riešenie:

Existuje konečnostavový zásobníkový automat A $= \{K,\ \Sigma,\ \Gamma,\ q_0,\ Z,\ F\}$ akceptujúci jazyk L. Keďže je jeho konštrukcia príliš veľká (cca. 40 stavov), a tým pádom aj rozsiahla funkcia δ, automat iba slovne opíšem.

V žiadnom prípade nie je tento automat najjednoduchší, existuje menejstavový automat, ktorý lepšie znázorňuje silu nedeterminizmu, na druhej strane je však nadmieru ťažké správnosť automatu dokázať, nieto ešte dôkaz pochopiť...

Automat bude mať 3 vetvy, podľa rozdielu dĺžok slov u, v v binárnom zápise:

$\boxed{||u| - |v|| \geq 2 :}$ Čísla, ktorých dĺžky binárnych zápisov sa líšia aspoň o dva, určite nie sú za sebou idúce. Nech $|u| = k, |v| = k+2$. Potom u je najväčšie vtedy, ak $u = \underbrace{11\ldots1}_{k}$ a v je najmenšie, ak $v = 1\underbrace{00\ldots0}_{k+1}$. Teda $u = 2^k - 1, v = 2^{k+1}$, čo zrejme nie sú nasledujúce čísla. Takže prvá vetva bude akceptovať práve takéto slová. (Pozn.: nasl. popis automatu bude stučný, kvôli prehľadnosti vynechávam detaily ako kontrola 'leading zeroes', atď) V jednom stave prečítam celé slovo u, pričom v zásobníku si zaznamenávam dĺžku tohto slova (pridávaním nejakého symbolu). Po prečítaní symbolu c začnem čítaním slova v symboly zo zásobníku vyhadzovať. Akceptujem v dvoch prípadoch:

1. Pokiaľ narazím na počiatočný zásobníkový symbol a viem prečítať ešte aspoň dva znaky slova v, presuniem sa do akceptačného stavu, kde dočítam zvyšok vstupu.
2. Ak som dočítal vstup a v zásobníku mám ešte aspoň dva znaky (okrem počiatočného), slovo tiež akceptujem.

$\boxed{||u|-|v||=1:}$ Jediný prípad, kedy nesmieme akceptovať takéto slovo je, ak $u = 1\dots1$ a $v = 10\dots0$. Automat v tejto vetve bude pracovať takto: Na začiatku sa nederministicky rozhodne, či sa $u = 1\dots1$ alebo nie:

1. Ak áno, v jednom stave (podobne ako v prvej vetve) zaznamená, koľko ich je. Prečíta c, 1 a zo zásobníku vyberiem jeden znak. Potom už len čítam nuly a vyprázdňujem zásobník. Ak na vstupe príde 1, podobne ako vo vetve 1 skontrolujem dĺžku, ak je OK tak akceptujem.

2. Ak nie, postupujem rovnako ako v predchádzajúcom kroku, až na to, že v slove u nepočítam len jednotky, ale aj nuly.

$\boxed{||u|-|v||=0:}$ Všimnime si najprv vlastnosť binárnych zápisov takýchto čísel. Ak $|u| = |v|$ a u je predchodca v, potom ak $u = w0\underbrace{1\dots1}_{k}, w \in 1\{0,1\}^*$, tak $v = w1\underbrace{0\dots0}_{k}$. Ak $u = 1\dots1$ a $|u| = |v|$, môžeme slovo akceptovať (opäť klasickou konštrukciou počítania symbolov pomocou zásobníku). Inak sa v podslove u nachádza aspoň jedna nula. Nedeterministicky si vyberieme jednu z nich. Môžu nastať opäť dva prípady:

1. Vybrali sme poslednú nulu podslova u. Pri čítaní ostatných jednotiek, budeme pridávať do zásobníku nejaký znak, až príde c. Pri slove v nedeterministicky odhadnem jeho poslednú jednotku. Prejdem do stavu, kde čítam nuly a vyprázdňujem zásobník, až kým sa nevyprázdni (a ešte som nedočítal pásku), alebo neskončí slovo (v zásobníku aspoň jeden symbol). Akceptujem.

2. Ak vybraná nula nie je posledná (t.j. načítali sme ešte aspoň jednu), zaznamenám koľká od konca bola. Prejdem cez znak c a vyberiem jednotku, ktorá je rovnako vzdialená od konca. Pokiaľ som takú našiel, slovo akceptujem.

Dôkaz:

Treba dokázať dve inklúzie: L(A)⊆L a L(A)⊇L.

$\boxed{\text{L(A)}\subseteq\text{L}}$ Musíme dokázať, že každé slovo akceptované automatom A, je aj z jazyku L. Ak $w \in$ L(A), potom muselo byť akceptované jedným zo sedem akceptačných stavov (2+2+3). Pre každý stav $q \in F$ dokážem, že pre každé slovo w akceptované stavom q platí: $w \in L$. (Podrobný dôkaz urobím len pre jeden stav, zvyšné stavy sa dokážu analogicky.) Nech w je slovo akceptované 3. vetvou prípadom 2 automatu A. Aby sa slovo w dostalo do koncového stavu, musí mať istý tvar. Podslovo u musí začínať jednotkou, za ktorou nasleduje niekoľko znakov 0,1 a nula, ktorú chceme skúmať. Keďže nie je v podslove u posledná, po niekoľkých jednotkách musíme načítať znovu nulu. Dočítame slovo u. T.j. $u = 1\alpha_1 01^n 0\alpha_2$, kde $n \geq 0$ a $\alpha_1, \alpha_2 \in \{0,1\}^*$. Nech $k = |1^n 0\alpha_2|$, potom kontrolujeme práve $(k+1)$-vý bit od konca. Podobne dočítame terminál

c a podslovo v, pričom $v = 1\alpha_3 1\alpha_4$, kde $\alpha_3, \alpha_4 \in \{0,1\}^*$ a $|\alpha_4| = k$, pretože akceptujeme len s počiatočným zásobníkovým symbolom. V predpoklade máme, že kontrolovaná nula, nebola posledná, teda zmena nuly na jednotku nebola dôsledkom 'carry'. Z toho vyplýva, že čísla u, v sa určite nelíšia o jedna, teda slovo w patrí do jazyka L.

$\boxed{\text{L(A)} \supseteq \text{L}}$ Pre každé slovo w $\in$ L, existuje v A akceptačný výpočet. Z opisu konštrukcie automatu vyplýva, že slovo bude akceptované aspoň v jednej vetve (podľa rozdielu dĺžok podslov u, v). Keďže sme pri konštrukcii rozobrali všetky možné tvary vstupného slova w, teda pokiaľ je z jazyka L, bude akceptované aj automatom A.