1.
Napište regulární výraz pro jazyk nad abecedou {0, 1},

a) jehož slova obsahují pouze nuly,
b) jehož každé slovo obsahuje právě jedinou jedničku,
c) jehož každé slovo obsahuje alespoň jednu jedničku,
d) jehož každé slovo obsahuje alespoň dvě jedničky,
e) jehož slova obsahují sudý počet jedniček,
f) jehož slova obsahují lichý počet jedniček.

Řešení

a) 00*
b) 0*10*
c) Libovolný řetězec z nul a jedniček lze popsat například výrazem (0*1*)*. Takový řetězec bude stát před i za jedničkou, o níž víme, že ve slově být musí: (0*1*)*1(0*1*)*. Podobně lze vytvořit výraz (0+1)*1(0+1)*, jenž pro menší počet svých symbolů se zdá být výhodnější. Postupem od začátku slova lze dojít k úvaze popsané slovy: „Nejprve libovolný počet nul, pak jednička, pak již cokoli“ t.j. 0*1(0*1*)* nebo 0*1(0+1)*. Stejně tak odzadu, t.j. (0*1*)*10* nebo (0+1)* 10*.
d) Zřejmě bude nejúspornější pokračovat jako v závěrečných případech předchozího řešení a vytvořit např. výraz 0*10*1(0+1)*.
e) Každé slovo obsahuje na začátku i na konci libovolný počet nul. Vnitřek musí obsahovat opakující se úseky (obecně různé délky!), z nichž každý obsahuje právě dvě jedničky: 0*(10*10*)*0*, Poslední opakující se nulu můžeme vypustit: 0*(10*10*)*. Podobně můžeme vytvořit výraz (0*10*1)*0. Stejně tak můžeme uvažovat, že slovo se skládá výlučně z úseků obsahujících dvě jedničky, takže hledaný výraz může být také 0*+ (0*10*10*)*.
f) Vyjdeme z předchozího řešení, doplníme podřetězec obsahující jedinou jedničku 0*(10*10*)*10*. Podobně lze upravit ostatní řešení.

2.
Nad abecedou {0, 1}, jsou dány dva jazyky L_1 a L_2. Slova L_1 jsou popsána výrazem 0*1*0*1*0*, slova L_2 jsou popsána výrazem (01+10)*.

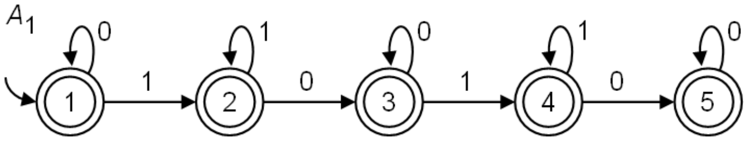
a) Najděte nejkratší neprázdné slovo v průniku $L_1 \cap L_2$.
b) Najděte nejdelší slovo v průniku $L_1 \cap L_2$.
c) Najděte nejkratší slovo, které leží v L_1, ale neleží v L_2.
d) Najděte nejkratší slovo, které leží v L_2, ale neleží v L_1.
e) Najděte nejkratší slovo, které neleží v $L_1 \cup L_2$.

Řešení

a) Zřejmě 01 i 10 leží v obou jazycích, L_2 neobsahuje kratší neprázdná slova.
b) V žádném slově L_2 nemohu stát bezprostředně za sebou tři stejné symboly, stejně tak, žádné slovo L_2 nemůže končit nebo začínat dvojicí stejných symbolů. Nejdelší slovo L_1, které vyhovuje těmto omezením, je 01100110 a je zřejmé, že také patří do L_2.
c) Stačí slovo jednotkové délky: 0 nebo 1.
d) Stačí tři jedničky oddělené navzájem nulami: 101010 nebo 010101
e) Podobně jako v předchozí úloze hledané slovo musí obsahovat alespoň tři jedničky oddělené nulami, aby neleželo v L_1. Navíc, protože jazyk L_2 obsahuje jen slova sudé délky, stačí, aby mělo lichou délku: 10101.

3.
Sestrojte konečné automaty A_1 resp. A_2 rozpoznávající jazyky L_1 resp. L_2.

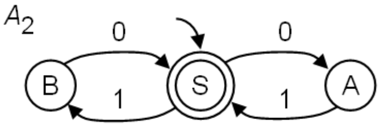
Intuitivní řešení pro (nedeterministický!) automat A_1:



Konstrukce A_1 metodou sousedů v regulárním výrazu:
Regulární výraz: 0*1*0*1*0*.
Označení symbolů: $0_1{}^*1_2{}^*0_3{}^*1_4{}^*0_5{}^*$.
Množina začátečních symbolů: $\{0_1, 1_2, 0_3, 1_4, 0_5\}$.
Množina (možných) sousedů: $\{0_10_1, 0_11_2, 0_10_3, 0_11_4, 0_10_5, 1_21_2, 1_20_3, 1_21_4, 1_20_5, 0_30_3, 0_31_4, 0_3,0_5,1_41_4, 1_40_5, 0_50_5\}$.
Množina koncových symbolů: $\{0_1, 1_2, 0_3, 1_4, 0_5\}$ (v tomto případě stejná jako množina začátečních symbolů).
Množina stavů A_1 : $\{q_0, 0_1, 1_2, 0_3, 1_4, 0_5\}$ (q_0 je počáteční stav).
Množina koncových stavů A_1 : $\{q_0, 0_1, 1_2, 0_3, 1_4, 0_5\}$.
Přechodové zobrazení nedeterministického automatu A_1 dané tabulkou:

	0	1
q_0	$0_1, 0_3, 0_5$	$1_2, 1_4$
0_1	$0_1, 0_3, 0_5$	$1_2, 1_4$
1_2	$0_3, 0_5$	$1_2, 1_4$
0_3	$0_3, 0_5$	1_4
1_4	0_5	1_4
0_5	0_5	

Intuitivní řešení pro (nedeterministický!) automat A_2:



Konstrukce A_2 metodou sousedů v regulárním výrazu:
Regulární výraz: (01+10)*.
Označení symbolů: $(0_11_2+1_30_4)^*$.
Množina začátečních symbolů: $\{0_1, 1_3\}$.
Množina (možných) sousedů: $\{0_11_2, 1_20_1, 1_21_3, 1_30_4, 0_40_1, 0_41_3\}$.
Množina koncových symbolů: $\{1_2, 0_4\}$.
Množina stavů A_1 : $\{q_0, 0_1, 1_2, 1_3, 0_4\}$ (q_0 je počáteční stav).
Množina koncových stavů A_1 : $\{q_0, 1_2, 0_4\}$.
Přechodové zobrazení nedeterministického automatu A_1 dané tabulkou:

	0	1
q_0	0_1	1_3
0_1		1_2
1_2	0_1	1_3
1_3	0_4	
0_4	0_1	1_3

4.

Gramatika G_4 generuje regulární jazyk L_4 nad abecedou {0, 1}. Jazyk L_4 představuje právě všechny binární zápisy nezáporných celých čísel dělitelných třemi (včetně prázdného slova).
G_4:

```
S → 0S | 1A | 0
A → 0B | 1S | 1
B → 0A | 1B
```

Pumping lemma pro jazyk L_4:
Existuje (dosud nevíme, jaké, ale nějaké pevné a konečné) číslo $p \geq 0$ s následující vlastností:
Každé slovo x jazyka L_4 o délce alespoň p obsahuje podřetězec v, takový, že zároveň se slovem x jazyk L_4 obsahuje všechna slova ve tvaru uv^kw ($k \geq 0$), a přitom $x = uvw$.
Neformálně to znamená, že když podřetězec v ve slově x libovolněkrát zopakujeme (těsně za sebou), dostaneme vždy zase slovo z L_4.
Převedeno do řeči dělitelnosti: V každém dostatečně dlouhém zápisu binárního čísla dělitelného 3 existuje podřetězec, jehož libovolným zopakováním dostaneme opět zápis čísla dělitelného třemi.

Například ve slově y = 1001110 představujícím binární zápis čísla 78, můžeme zvolit podřetězec 11. Přesvědčte se, že všechna slova ve tvaru 1001(11)*0 představují zápis binárního čísla dělitelného 3, a tudíž patří do L_4.

Určete hodnotu p pro jazyk L_4.

Řešení

Všechna slova ve tvaru 1001(11)*0 představují zápis binárního čísla dělitelného 3.
Zdůvodnění: Nula na konci nehraje roli, číslo 1001(11)*0 je dělitelné třemi právě tehdy, když je dělitelné třemi číslo 1001(11)*. Prefix 1001 má hodnotu 9, ta je dělitelná třemi. Pro $k \geq 0$ označme $y_k = 1001(11)^k$. Platí: hodnota y_{k+1} = hodnota y_k násobená 4 (posun o dvě číslice doleva) a zvětšená o 3 (hodnota připojeného řetězce 11). Nyní ze schématu úplné indukce přímo plyne, že hodnota y_k je dělitelná třemi pro každé $k \geq 0$, tedy $1001(11)^*0 \in L_4$.

p = 0.
V každém slově L_4. lze položit $u = w = \varepsilon$. Zřetězení téhož slova představuje opět číslo dělitelné třemi (důvod je podobný jako výše).

Naposled přeformulujeme pumping lemma pro L_4:
Každé slovo x jazyka L_4 o délce alespoň 0 obsahuje podřetězec v, takový, že zároveň se slovem x jazyk L_4 obsahuje všechna slova ve tvaru uv^kw ($k \geq 0$), a přitom (v případě jazyka L_4:) $x = uvw$, $u = w = \varepsilon$, t.j. $x = v$. To znamená, že když x libovolněkrát za sebou zřetězíme, dostaneme vždy zase slovo z L_4.
Převedeno do řeči dělitelnosti: Každý zápis binárního čísla dělitelného lze vícekrát za sebou zřetězit a opět dostaneme zápis čísla dělitelného třemi.