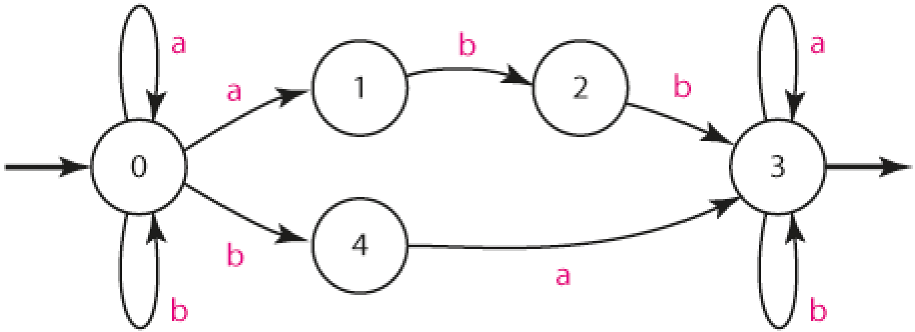
Využití NKA při návrhu syntaktického analyzátoru

Příklad 23

Navrhněte NKA, který akceptuje všechny řetězce obsahující -abb- nebo -ba-

Tento automat bychom jistě dokázali navrhnout jako deterministický. Příklad má na jednodu-ché úloze ukázat, že návrh NKA může být jednodušší; absence „nápadu a intuice“ přitom bude kom-penzována mechanickou činností při vytváření deterministického ekvivalentu.

Příklad 23



	a	b
→ 0	{0,1}	{0,4}
1	-	{2}
2	-	{3}
← 3	{3}	{3}
4	{3}	-

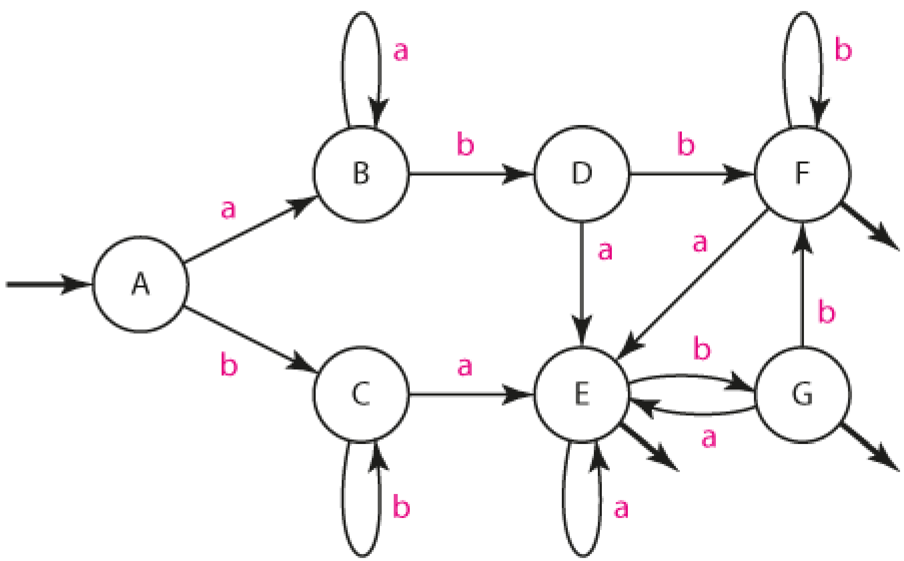
Příklad 23

	a	b
→ 0	{0,1}	{0,4}
1	-	{2}
2	-	{3}
← 3	{3}	{3}
4	{3}	-

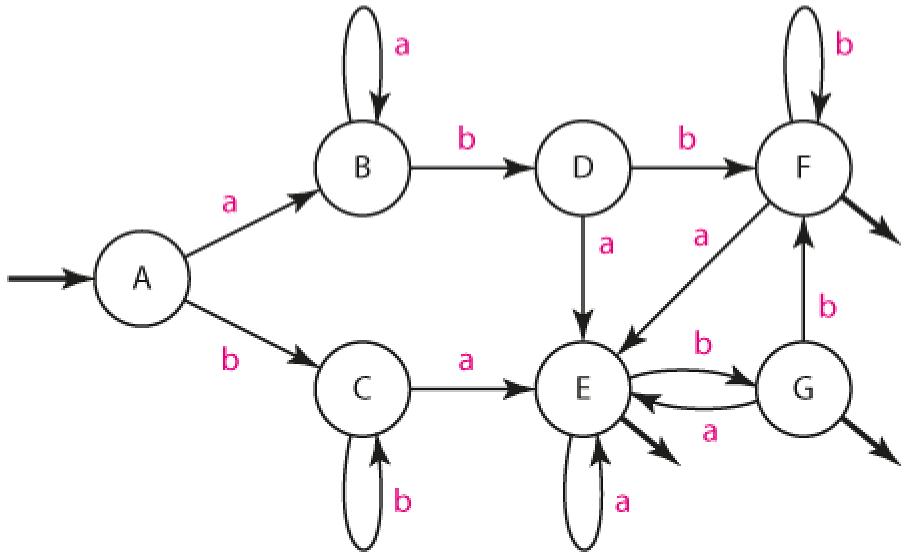
	a	b
→ 0	{0,1}	{0,4}
{0,1}	{0,1}	{0,2,4}
{0,4}	{0,1,3}	{0,4}
{0,2,4}	{0,1,3}	{0,3,4}
← **{0,1,3}**	{0,1,3}	{0,2,3,4}
← **{0,3,4}**	{0,1,3}	{0,3,4}
← **{0,2,3,4}**	{0,1,3}	{0,3,4}

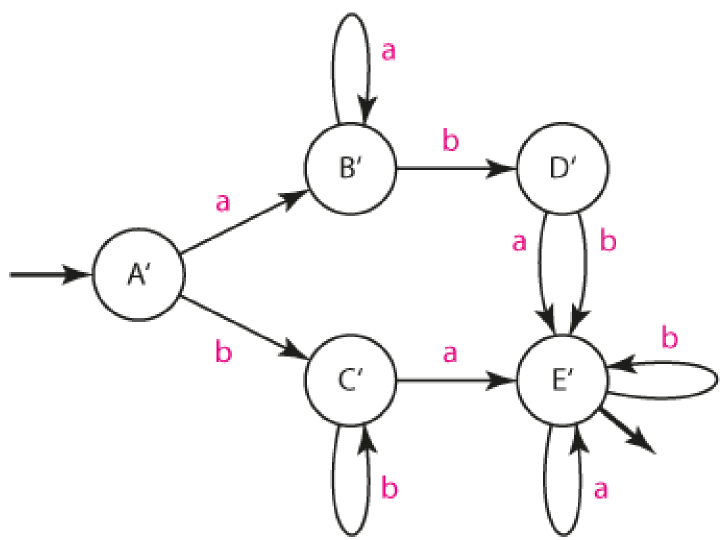
Příklad 23

		a		b	
A	→ 0	B	{0,1}	C	{0,4}
B	{0,1}	B	{0,1}	D	{0,2,4}
C	{0,4}	E	{0,1,3}	C	{0,4}
D	{0,2,4}	E	{0,1,3}	F	{0,3,4}
E	← {0,1,3}	E	{0,1,3}	G	{0,2,3,4}
F	← {0,3,4}	E	{0,1,3}	F	{0,3,4}
G	← {0,2,3,4}	E	{0,1,3}	F	{0,3,4}



Příklad 23





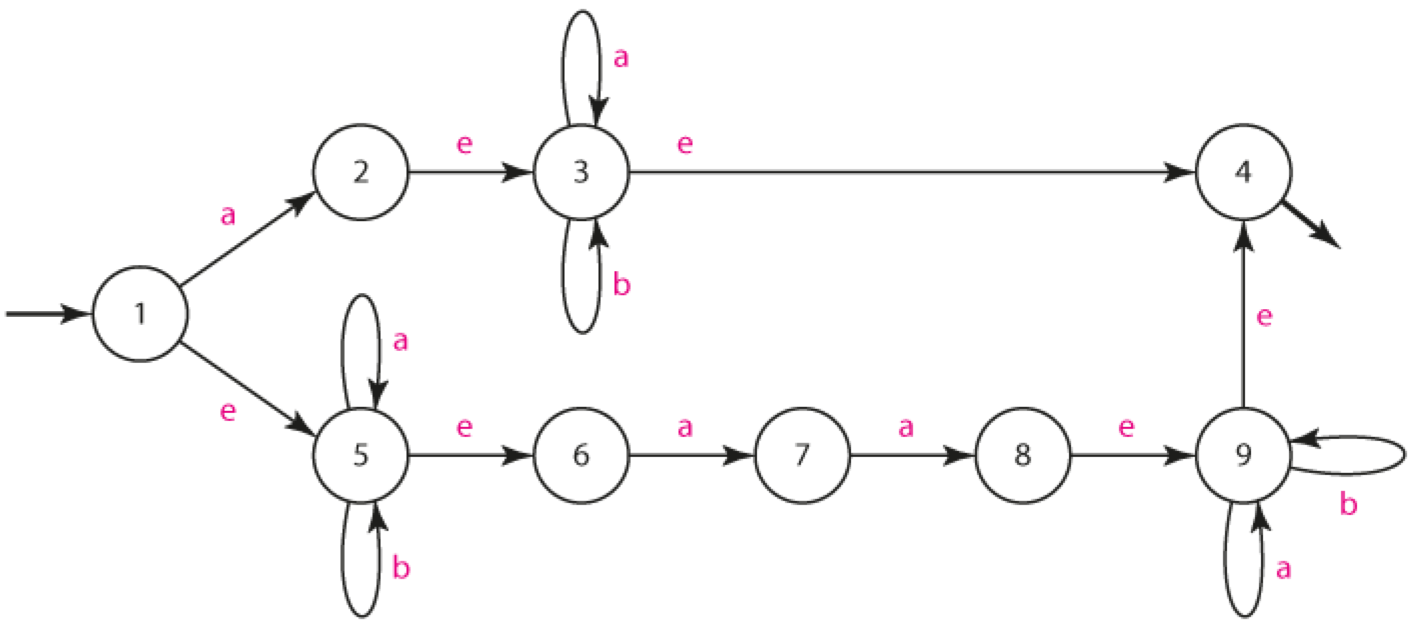
Příklad 24

Navrhněte NKA, který akceptuje řetězce popsané regulárním výrazem $a(a+b)^* + (a+b)^*aa(a+b)^*$.

Postupným rozkladem regulárního výrazu dojdeme k NKA.

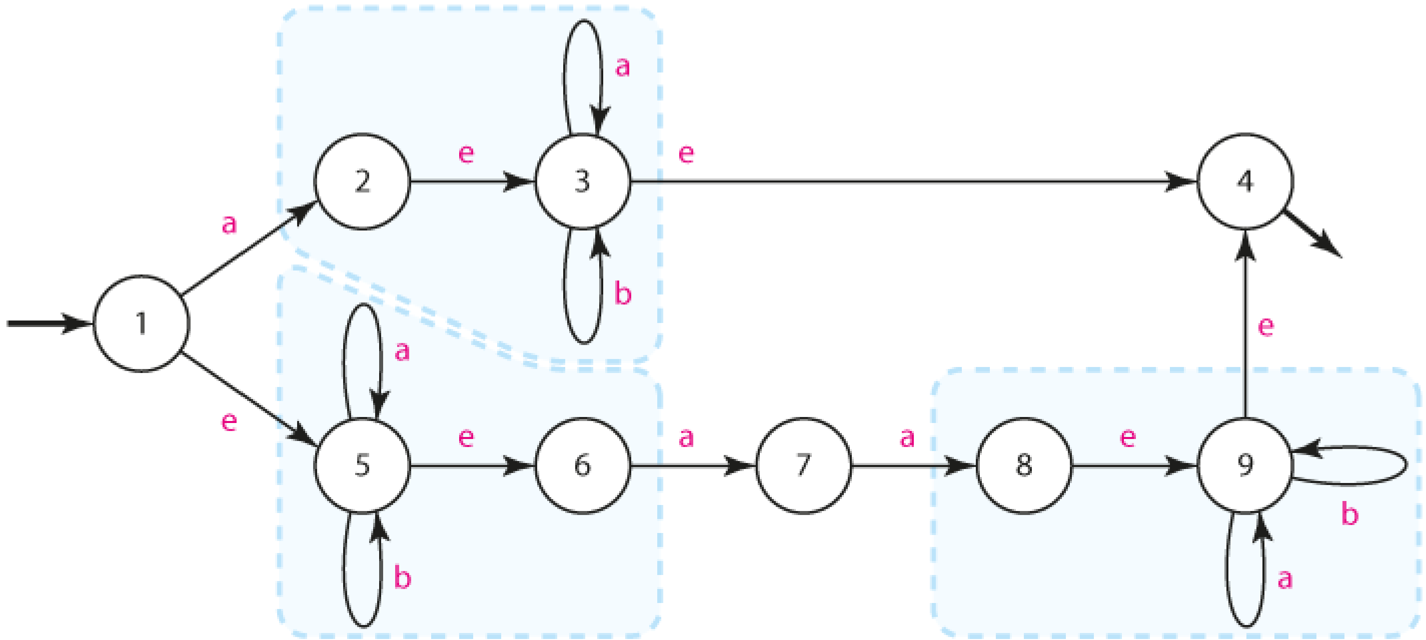
Příklad 24

a(a+b)* + (a+b)*aa(a+b)*



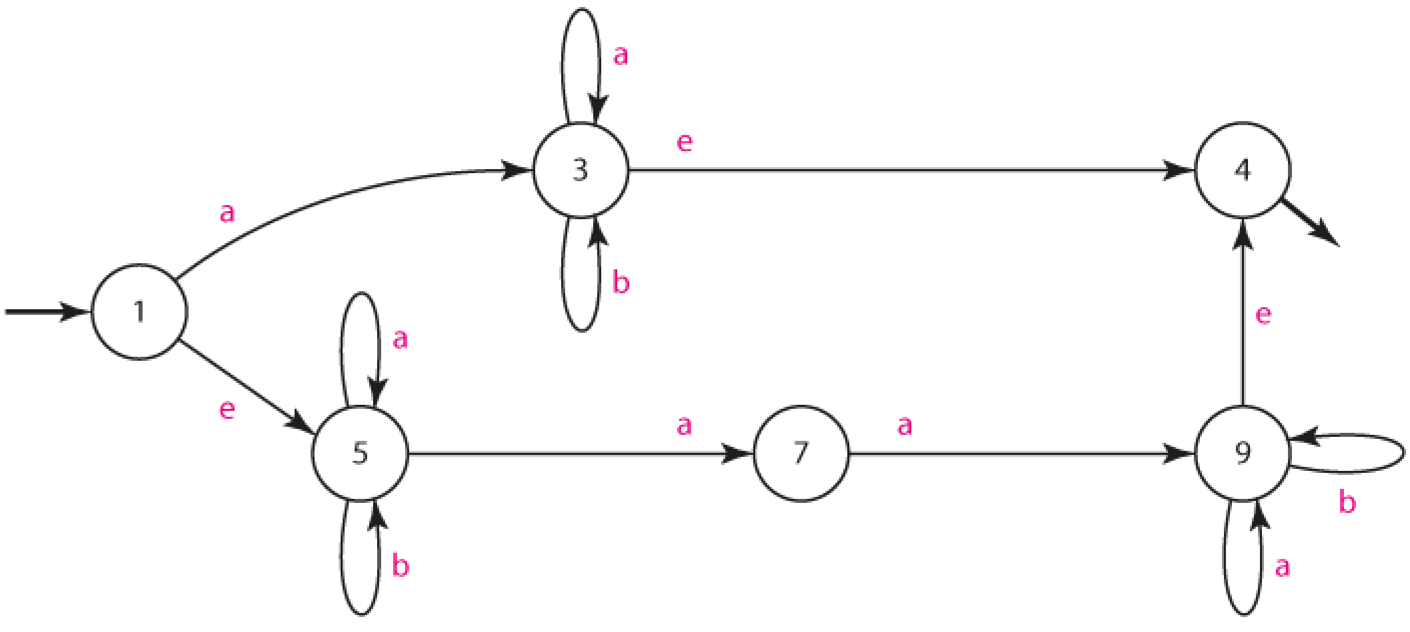
Příklad 24

V přechodovém grafu lze některé e-hrany vynechat a stavy sloučit.



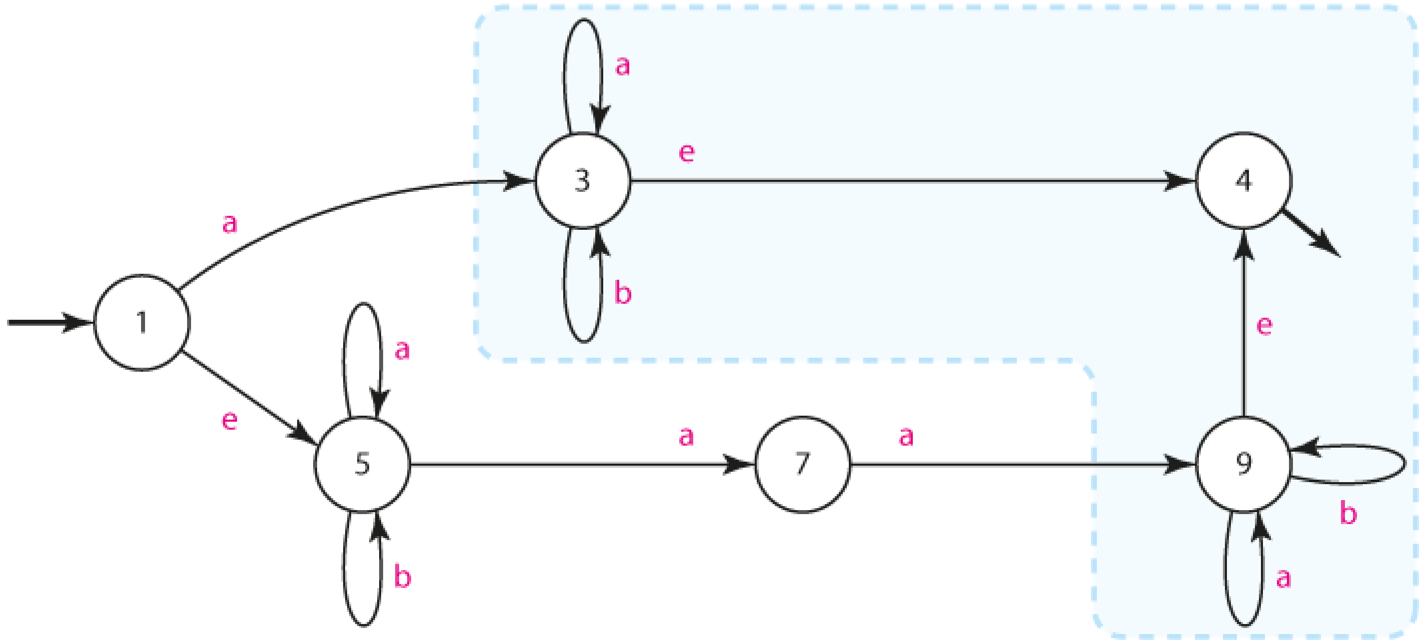
Příklad 24

Po vynechání e-hran tam, kde to jde:



Příklad 24

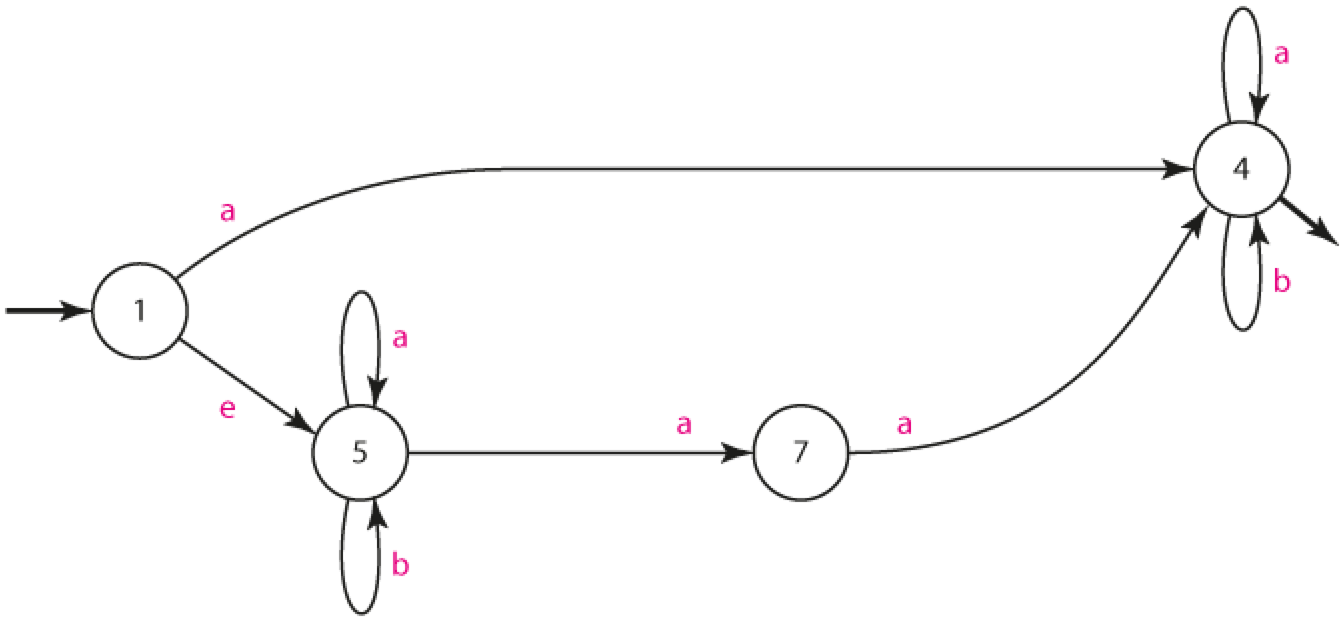
Další zjednodušení (intuitivní)



V nově vyznačené skupině automat setrvává jak vstupem *a*, tak vstupem *b*, a díky e-hranám je jedno, zda je koncový pouze stav 4, nebo i stavy 3 a 9.

Příklad 24

Po úpravě



Příklad 24

	a	b
→ 1	{4}	-
← 4	{4}	{4}
5	{5,7}	{5}
7	{4}	-

e-následníci
{1,5}
{4}
{5}
{7}

	a	b
→ {1,5}	{4,5,7}	{5}
← {4,5,7}	{4,5,7}	{4,5}
{5}	{5,7}	{5}
← {4,5}	{4,5,7}	{4,5}
← {5,7}	{4,5,7}	{5}

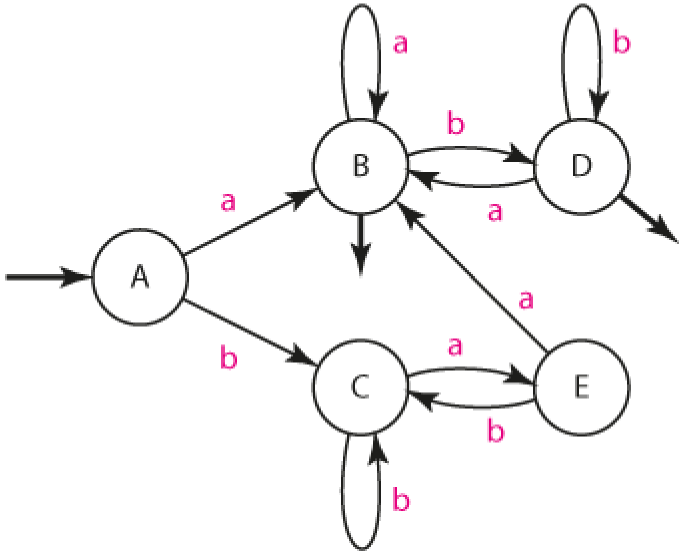
Příklad 24

	a	b
→ {1,5}	{4,5,7}	{5}
← {4,5,7}	{4,5,7}	{4,5}
{5}	{5,7}	{5}
← {4,5}	{4,5,7}	{4,5}
{5,7}	{4,5,7}	{5}

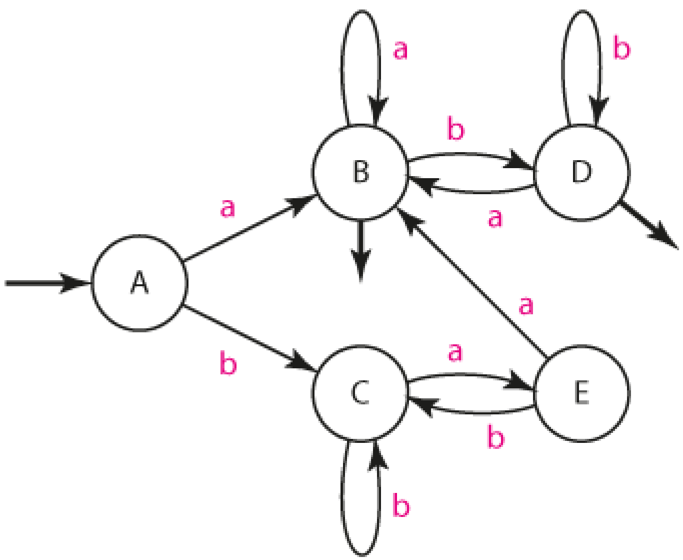
	a	b
→ A {1,5}	B {4,5,7}	C {5}
← B {4,5,7}	B {4,5,7}	D {4,5}
C {5}	E {5,7}	C {5}
← D {4,5}	B {4,5,7}	D {4,5}
E {5,7}	B {4,5,7}	C {5}

Příklad 24

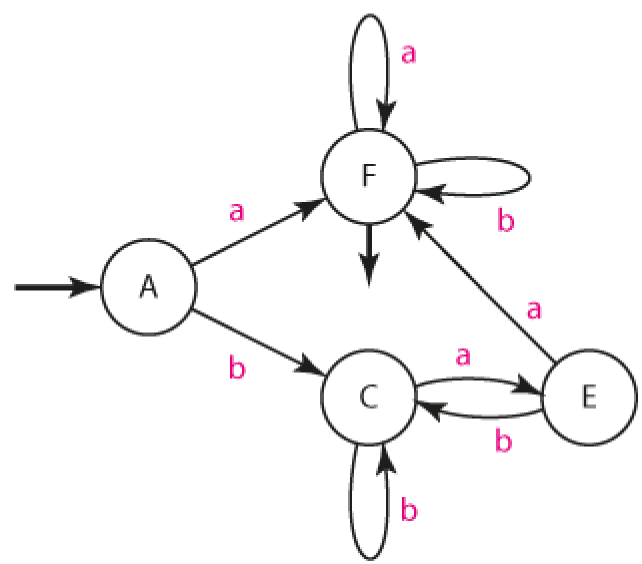
	a	b
→ **A** {1,5}	B {4,5,7}	C {5}
← **B** {4,5,7}	B {4,5,7}	D {4,5}
C {5}	E {5,7}	C {5}
← **D** {4,5}	B {4,5,7}	D {4,5}
E {5,7}	B {4,5,7}	C {5}



Příklad 24



Stavy B a D jsou ekvivalentní, lze ponechat jen jeden (přejmenován na F)

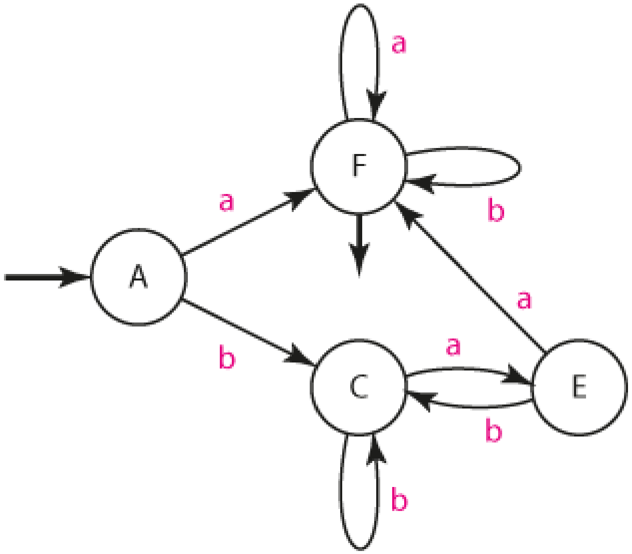


Příklad 24

$$a(a+b)^* + (a+b)^*aa(a+b)^*$$

Jaké řetězce jsou „ty správné“?

Začínají písmenem *a* a/nebo obsahují podřetězec *aa* .



Elementární příklady na teorii informace

Příklad 25

Zdroj zpráv produkuje znaky ze souboru $x=\{x_1,x_2,x_3,x_4\}$, přičemž pravděpodobnosti vyslání jednotlivých znaků jsou

$p(x_1)=0.2$, $p(x_2)=0.3$, $p(x_3)=0.4$, $p(x_4)=0.1$

Vypočtěte střední entropii zdroje, určete jeho redundanci.

Příklad 25

$p(x_1)=0.2$, $p(x_2)=0.3$, $p(x_3)=0.4$, $p(x_4)=0.1$

$$H(X) = -\sum_{i=1}^{4} p(x_i)\log_2 p(x_i) = -\frac{1}{\log_{10} 2}\cdot\sum_{i=1}^{4} p(x_i)\log_{10} p(x_i)$$

$$H(X) = -\frac{1}{log\ 2}\cdot(0.2\cdot\log 0.2 + 0.3\cdot\log 0.3 + 0.4\cdot\log 0.4 + 0.1\cdot\log 0.1\ ) =$$

$$= -\frac{1}{0.3010}\cdot(-\ 0.1398 - 0.1568 - 0.1592 - 0.1000\ ) = -\frac{1}{0.3010}\cdot(-0.5558) =$$

$=$ **1.8465 [bit]**

Příklad 25

Pro střední neurčitost zdroje vždy platí:

$0 \leq H(X) \leq \log_2 n$, kde n je počet písmen zdroje

Mezní hodnoty:

$H(X) = 0$ když má jedno písmeno p-st 1

$H(X) = \log_2 n$ když mají všechna písmena stejnou p-st

$p(x_1)=0.2$, $p(x_2)=0.3$, $p(x_3)=0.4$, $p(x_4)=0.1$ $0 \leq H(X) \leq \log_2 4 = 2$

$H(X) = $ **1.8465 [bit]** je „něco mezi“ (rozložení p-stí není rovnoměrné)

Příklad 25

Redundance (nadbytečnost) zdroje:

$$\rho = 1 - \frac{\mathrm{H(X)}}{\max \mathrm{H(X)}} = 1 - \frac{\mathrm{H(X)}}{\log_2 \mathrm{n}}$$

$p(x_1)=0.2$, $p(x_2)=0.3$, $p(x_3)=0.4$, $p(x_4)=0.1$ $\quad \mathrm{H(X)} = \mathbf{1.8465}$ **[bit]**

$$\rho = 1 - \frac{\mathrm{H(X)}}{\log_2 \mathrm{n}} = 1 - \frac{1.847}{\log_2 4} = 1 - \frac{1.847}{2} = \mathbf{0.0765}$$

Redundance je bezrozměrná veličina, lze ji vyjádřit v procentech.

Příklad 26

10000 šesticiferných čísel v desítkové soustavě

Kolik informace tabulka obsahuje?

Za předpokladu stejné p-sti všech číslic je neur-
čitost v každé desítkové cifře tabulky

$$H(X) = \log_2 10 \text{ [bit]}$$

Každá cifra tedy nese informaci

$$I(X) = H(X) = \log_2 10 \text{ [bit]}$$

Příklad 26

10000 šesticiferných čísel v desítkové soustavě

Každá cifra tedy nese informaci

$$I(X) = H(X) = \log_2 10 \text{ [bit]}$$

Jednotlivé číslice v tabulce považujeme za nezávislé $\Rightarrow$ informaci, kterou poskytují, můžeme sčítat. Číslic je v tabulce $n = 60.000$

Informace obsažená v tabulce:

$$I(TAB) = 60\,000 \cdot I(X) = 60\,000 \cdot \log_2 10 = 199\,315.69 \text{ [bit]}$$