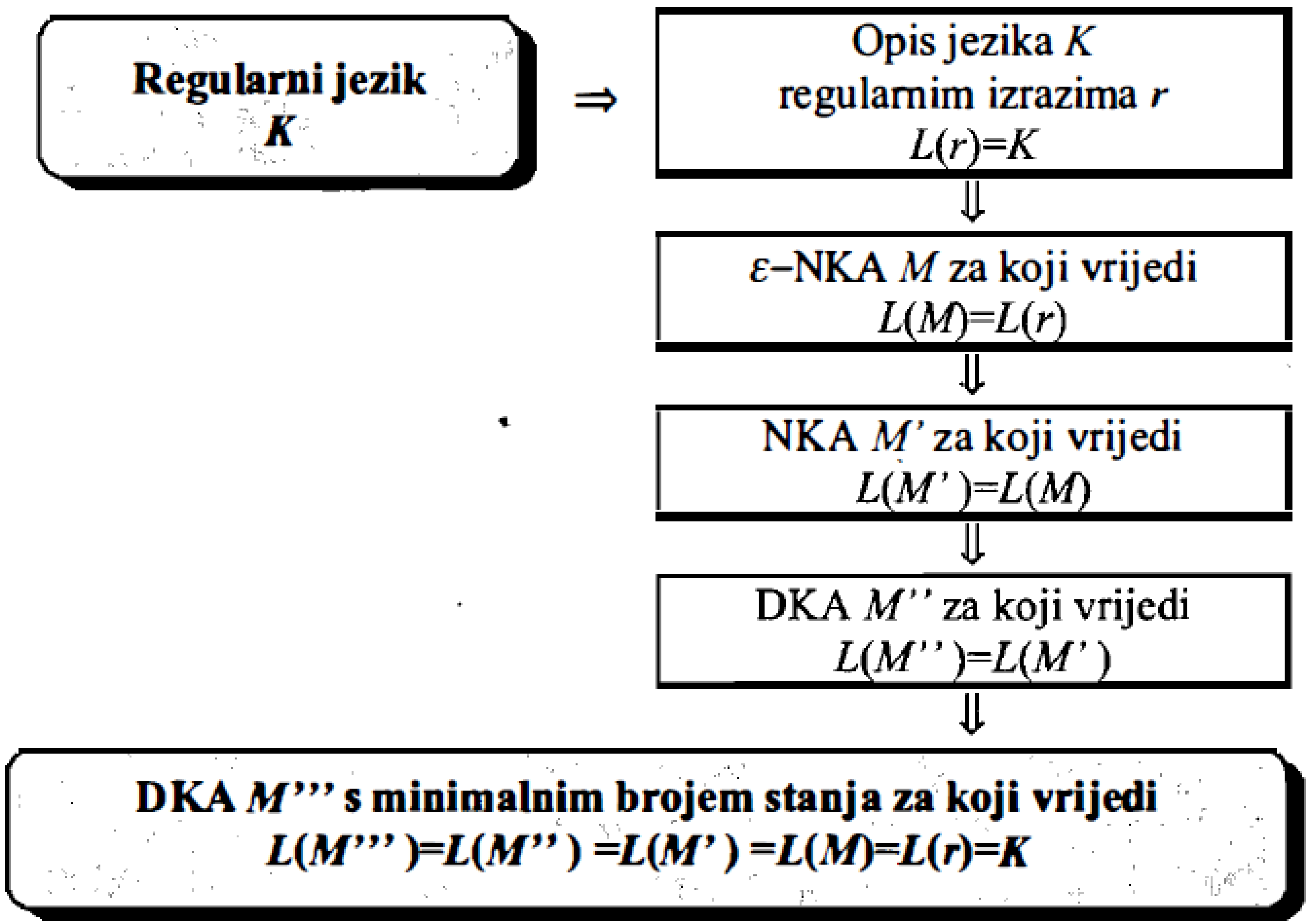
Regularni izrazi

Ako je neki jezik moguće opisati regularnim izrazima, onda je taj jezik sigurno regularan. Neka $L(r)$ označava jezik definiran regularnim izrazima r. Za bilo koji jezik $L(r)$ zadan regularnim izrazima r moguće je izgraditi DKA M za koji vrijedi $L(M)=L(r)$.

Postoji algoritam za pretvorbu regularnih izraza u ε−NKA. Na slici 2.34 prikazan je postupak gradnje DKA s minimalnim brojem stanja na temelju zadanih regularnih izraza.



Slika 2.34: Konstrukcija DKA s minimalnim brojem stanja za regularni jezik K

$r+s = s+r$	+ jest komutativno
$r+(s+t) = (r+s)+t$	+ jest asocijativno
$(rs)t = r(st)$	nadovezivanje jest asocijativno
$r(s+t) = rs+rt$ $(s+t)r = sr+tr$	distributivnost nadovezivanja nad +
$\varepsilon r = r\varepsilon = r$	ε jest neutralni element za nadovezivanje
$r^* = (r+\varepsilon)^*$	relacija između + i *
$r^{**} = r^*$	idempotentnost

Tablica 2.5: Algebarski zakoni koji vrijede za regularne izraze (r, s i t su regularni izrazi)

Primjer 2.7. Za binarnu abecedu $\Sigma=\{0, 1\}$ moguće je dati sljedeće primjere regularnih izraza i jezika:

1) Regularni izraz 01 definira jezik: $L(01)=\{01\}$.

2) Regularni izraz 0+1 definira jezik: $L(0+1)=\{0, 1\}$.

3) Regularni izraz (0+1)(0+1) definira jezik: $L((0+1)(0+1))=\{00, 01, 10, 11\}$.

4) Regularni izraz 1* definira jezik u kojem su prazni niz i svi nizovi koji se sastoje od proizvoljnog broja znakova 1: $L(1^*)=\{\varepsilon, 1, 11, 111, ..., 1111111, ...\}$.

5) Regularni izraz (0+1)* definira jezik u kojem su prazni niz i svi nizovi znakova 0 i 1: $L(0+1)^*=\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, ..., 0111101, ...\}$.

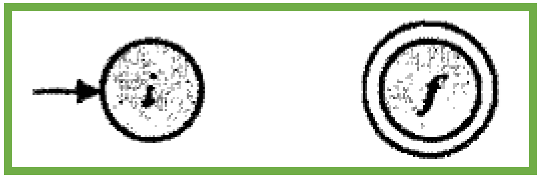
6) Regularni izraz (0+1)*00(0+1)* definira jezik u kojem bilo koji niz barem na jednom mjestu ima najmanje dva uzastopna znaka 0: $L((0+1)^*00(0+1)^*)=\{00, 000, 100, 001, 0000, 0100, 1000, 1100, 0001, 0010, 0011, ..., 0100111001, ...\}$.

7) Regularni izraz 0*1* definira jezik u kojem su prazni niz i nizovi u kojima iza proizvoljnog broja znakova 0 slijedi proizvoljni broj znakova 1: $L(0^*1^*)=\{\varepsilon, 0, 1, 00, 01, 000, 001, 011, 111, ..., 000111111, ...\}$.

Konstrukcija ε-NKA na temelju zadanih regularnih izraza

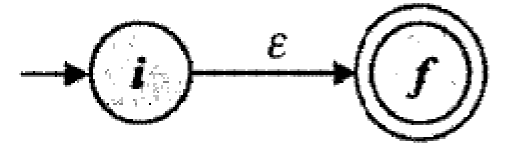
Za bilo koji regularni izraz r moguće je izgraditi ε–NKA M tako da vrijedi $L(M)=L(r)$:

$p1)$ Za regularni izraz $\boxed{\emptyset}$ koji definira jezik $L(\emptyset)=\{\}$ konstruira se ε–NKA $M=(\{i, f\}, \Sigma, \{\}, i, \{f\})$. Početno stanje jest i, a prihvatljivo stanje jest $f \in F$. Dijagram stanja izgrađenog ε–NKA je:



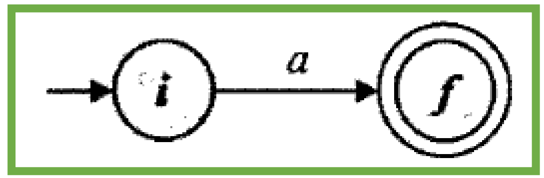
Napomena. Za bilo koji $b \in (\Sigma \cup \{\varepsilon\})$, skup $\delta(i, b)$ jest prazan skup. Ne postoji niti jedan prijelaz u prihvatljivo stanje f. Budući da je početno stanje i neprihvatljivo, ε–NKA M ne prihvaća niti jedan niz, uključujući i prazni niz ε.

$p2)$ Za regularni izraz $\boxed{\varepsilon}$ koji definira jezik $L(\varepsilon)=\{\varepsilon\}$ konstruira se ε–NKA $M=(\{i, f\}, \Sigma, \{\delta(i, \varepsilon)=f\}, i, \{f\})$. Početno stanje jest i, a prihvatljivo stanje jest $f \in F$. Dijagram stanja izgrađenog ε–NKA je:



Napomena. Za bilo koji $b \in \Sigma$, skup $\delta(f, b)$ jest prazan skup. Prijelaz $\delta(i, \varepsilon)=\{f\}$ omogućuje prihvaćanje praznog niza. ε–NKA M ne prihvaća niti jedan niz, osim praznog niza ε.

*p*3) Za regularni izraz a koji definira jezik $L(a)=\{a\}$ konstruira se ε–NKA $M=(\{i, f\}, \Sigma, \{\delta(i, a)=f\}, i, \{f\})$. Početno stanje jest i, a prihvatljivo stanje jest $f \in F$. Dijagram stanja izgrađenog ε–NKA je:

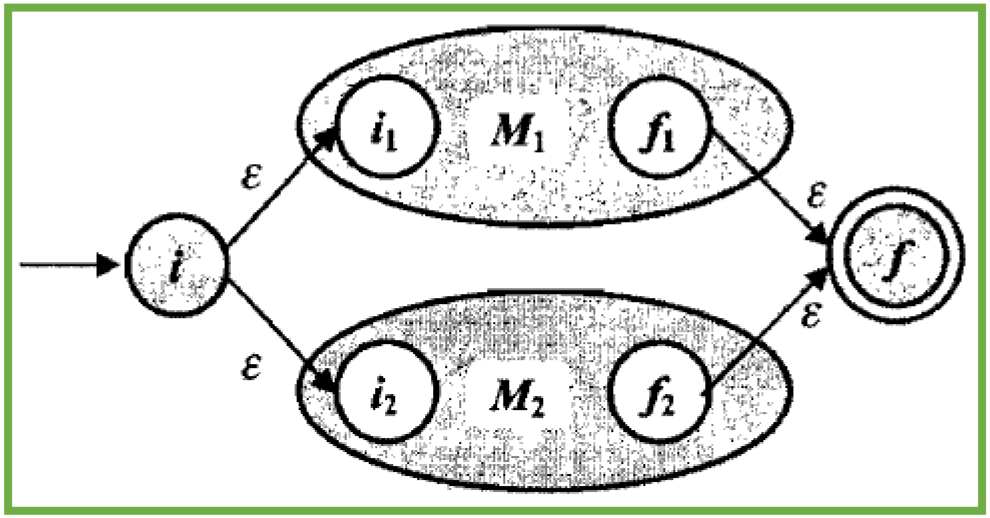


Napomena. Za bilo koji $b \in (\Sigma \cup \{\varepsilon\})$ za koji vrijedi $b \neq a$, skup $\delta(f, b)$ jest prazan skup. Prijelaz $\delta(i, a)=\{f\}$ omogućuje prihvaćanje niza a. ε–NKA M ne prihvaća niti jedan niz, osim niza a. Ne prihvaća se ni prazni niz ε.

$p4$) Za regularni izraz $\boxed{r_1+r_2}$ koji definira jezik $L(r_1+r_2)=L(r_1)\cup L(r_2)$ konstruira se ε–NKA M na sljedeći način. Pretpostavimo da su prethodno izgrađeni ε–NKA $M_1=(Q_1, \Sigma_1, \delta_1, i_1, \{f_1\})$ i $M_2=(Q_2, \Sigma_2, \delta_2, i_2, \{f_2\})$ takvi da vrijedi $L(M_1)=L(r_1)$ i $L(M_2)=L(r_2)$ i da nema prijelaza iz stanja f_1 i f_2 niti za jedan ulazni znak (tj. $\delta_1(f_1, a)=\emptyset$, $\forall a\in(\Sigma_1\cup\{\varepsilon\})$ i $\delta_2(f_2, b)=\emptyset$, $\forall b\in(\Sigma_2\cup\{\varepsilon\})$). Promjenom imena stanja u Q_1 i Q_2 postiže se da je $Q_1\cap Q_2=\{\}$. Konstruira se ε–NKA $M=(Q_1\cup Q_2\cup\{i, f\}, \Sigma_1\cup\Sigma_2, \delta, i, \{f\})$. Novo početno stanje jest i, a novo prihvatljivo stanje jest f. Stanja i_1 i i_2 nisu više početna stanja, te stanja f_1 i f_2 nisu više prihvatljiva. Funkcija δ određuje se na sljedeći način:

a) $\delta(i, \varepsilon)=\{i_1, i_2\}$,
b) $\delta(q, a)=\delta_1(q, a)$, $\forall q\in(Q_1- \{f_1\})$ i $\forall a\in(\Sigma_1\cup\{\varepsilon\})$,
c) $\delta(q, b)=\delta_2(q, b)$, $\forall q\in(Q_2- \{f_2\})$ i $\forall b\in(\Sigma_2\cup\{\varepsilon\})$,
d) $\delta(f_1, \varepsilon)=\delta(f_2, \varepsilon)=\{f\}$

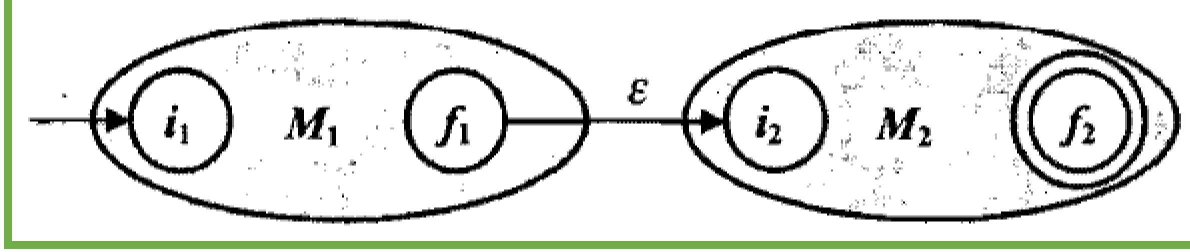
Dijagram stanja izgrađenog ε–NKA M koji prihvaća jezik $L(M)=L(r_1+r_2)$ je:



$p5)$ Za regularni izraz $\boxed{r_1 r_2}$ koji definira jezik $L(r_1r_2)=L(r_1)L(r_2)$ konstruira se ε–NKA M na sljedeći način. Pretpostavimo da su prethodno izgrađeni ε–NKA $M_1=(Q_1, \Sigma_1, \delta_1, i_1, \{f_1\})$ i $M_2=(Q_2, \Sigma_2, \delta_2, i_2, \{f_2\})$ takvi da vrijedi $L(M_1)=L(r_1)$ i $L(M_2)=L(r_2)$ i da nema prijelaza iz stanja f_1 i f_2 niti za jedan ulazni znak (tj. $\delta_1(f_1, a)=\emptyset$, $\forall a\in(\Sigma_1\cup\{\varepsilon\})$ i $\delta_2(f_2, b)=\emptyset$, $\forall b\in(\Sigma_2\cup\{\varepsilon\})$). Promjenom imena stanja u Q_1 i Q_2 postiže se da je $Q_1\cap Q_2=\{\}$. Konstruira se ε–NKA $M=(Q_1\cup Q_2, \Sigma_1\cup\Sigma_2, \delta, i_1, \{f_2\})$. Novo početno stanje jest i_1, a novo prihvatljivo stanje jest f_2. Stanje i_2 nije više početno stanje, te stanje f_1 nije više prihvatljivo. Funkcija δ određuje se na sljedeći način:

a) $\delta(q, a)=\delta_1(q, a)$, $\forall q\in(Q_1- \{f_1\})$ i $\forall a\in(\Sigma_1\cup\{\varepsilon\})$,
b) $\delta(q, b)=\delta_2(q, b)$, $\forall q\in Q_2$ i $\forall b\in(\Sigma_2\cup\{\varepsilon\})$,
c) $\delta(f_1, \varepsilon)=\{i_2\}$

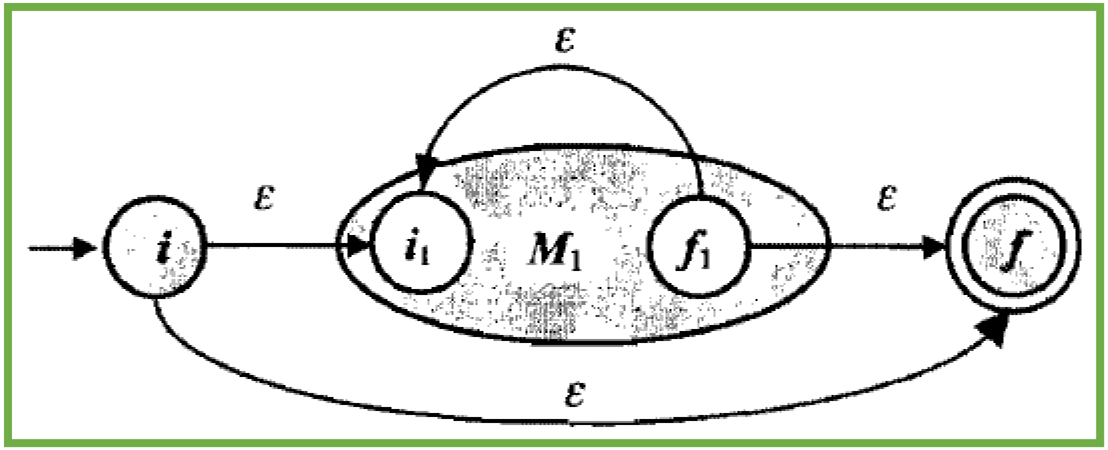
Dijagram stanja izgrađenog ε–NKA M koji prihvaća jezik $L(M)=L(r_1r_2)$ je:



*p*6) Za regularni izraz $\boxed{r_1*}$ koji definira jezik $L(r_1*)=L(r_1)*$ konstruira se ε–NKA M na sljedeći način. Pretpostavimo da je prethodno izgrađeni ε–NKA $M_1=(Q_1, \Sigma_1, \delta_1, i_1, \{f_1\})$ za koji vrijedi $L(M_1)=L(r_1)$ i nema prijelaza iz stanja f_1 niti za jedan ulazni znak (tj. $\delta_1(f_1, a)=\emptyset$, $\forall a\in(\Sigma_1\cup\{\varepsilon\})$). Konstruira se ε–NKA $M=(Q_1\cup\{i, f\}, \Sigma_1, \delta, i, \{f\})$. Novo početno stanje jest i, a novo prihvatljivo stanje jest f. Stanje i_1 nije više početno stanje, te stanje f_1 nije više prihvatljivo. Funkcija δ određuje se na sljedeći način:

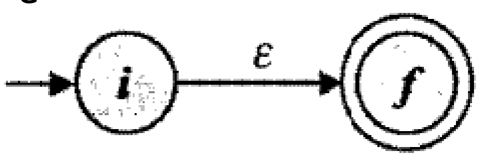
a) $\delta(i, \varepsilon)=\delta(f_1, \varepsilon)=\{i_1, f\}$,
b) $\delta(q, a)=\delta_1(q, a)$, $\forall q\in(Q_1- \{f_1\})$ i $\forall a\in(\Sigma_1\cup\{\varepsilon\})$,

Dijagram stanja izgrađenog ε–NKA M koji prihvaća jezik $L(M)=L(r_1*)$ je:

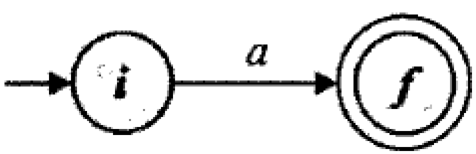


∅

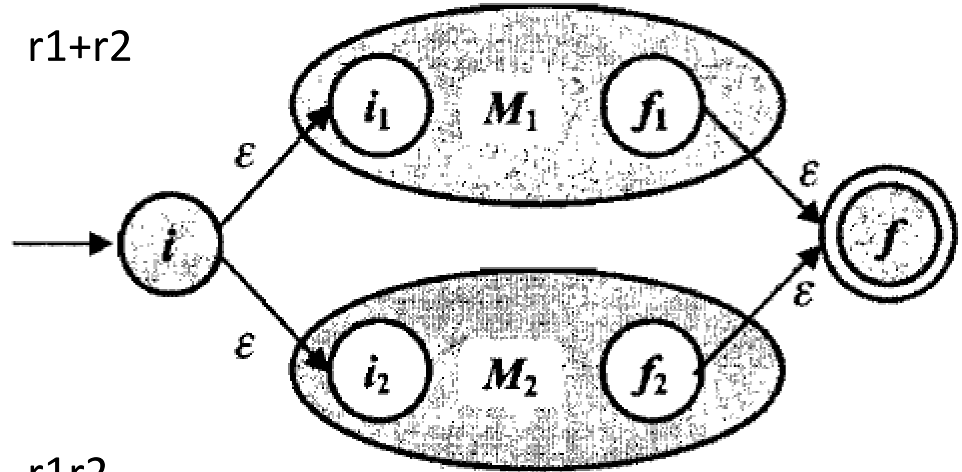
ε



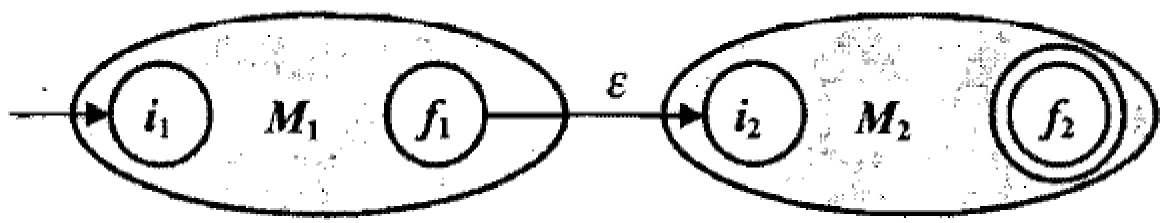
a



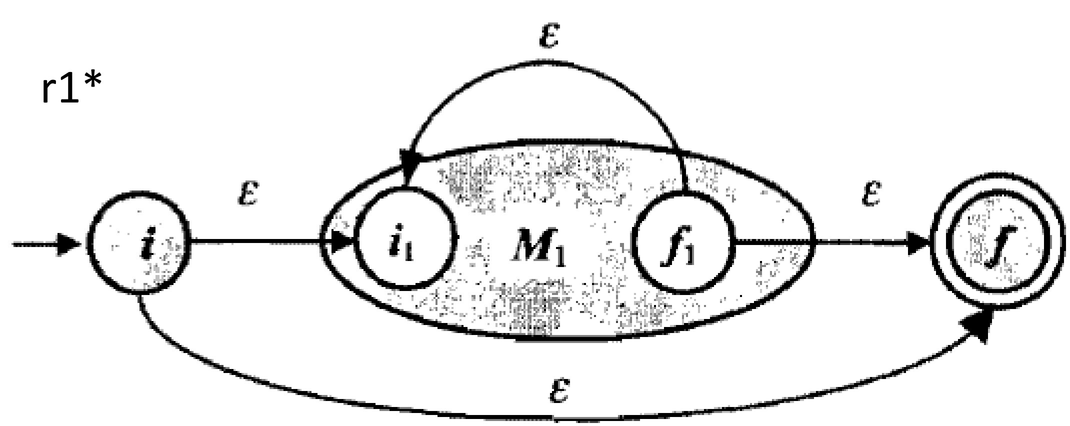
r1+r2



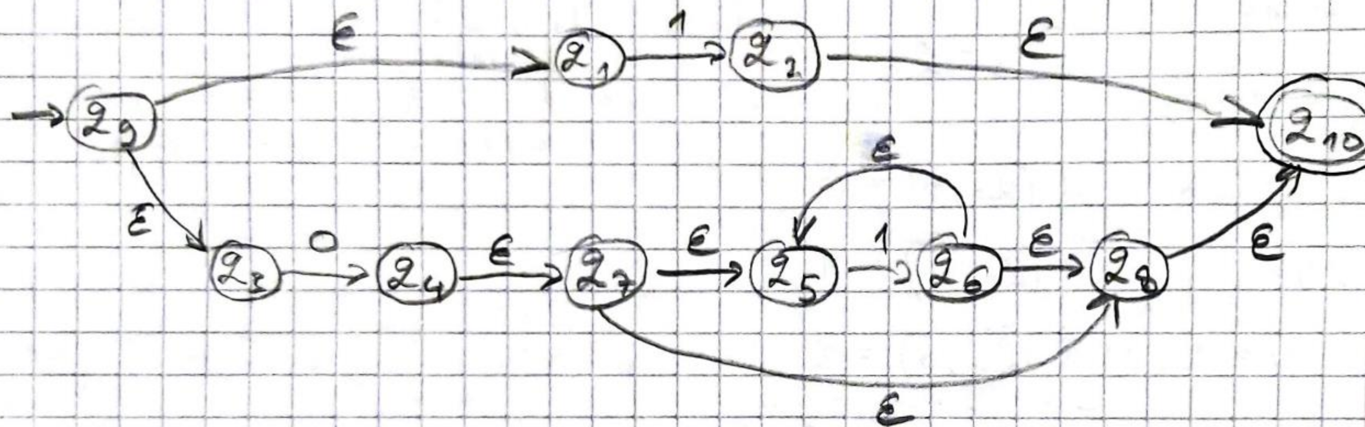
r1r2



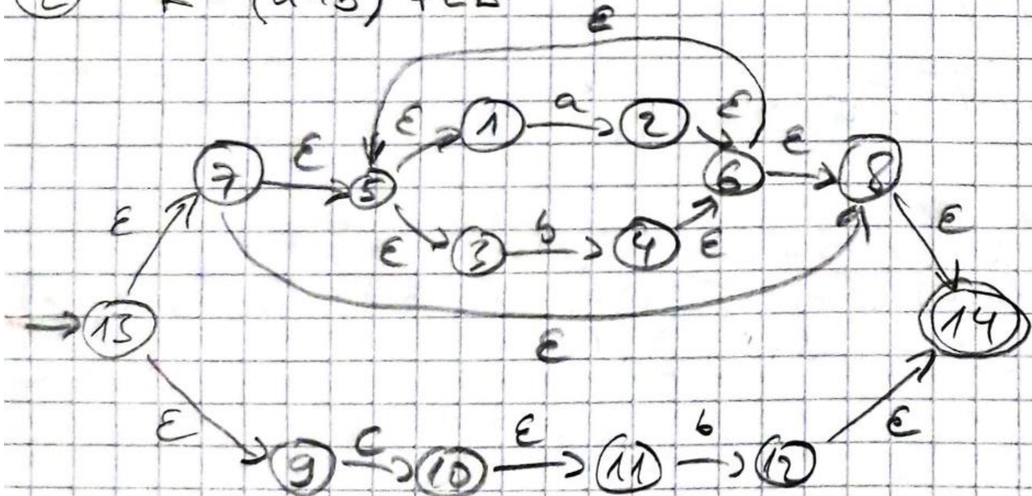
r1*



1 $R = 1+01^*$



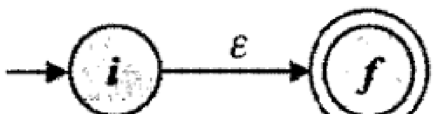
2 $R = (a+b)^*+cb$



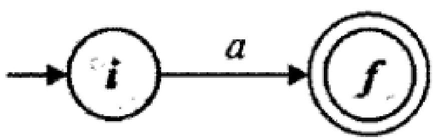
(3) R = ab (4) R = a*b (5) R = a+b (6) R = (ab)*+c

Ø

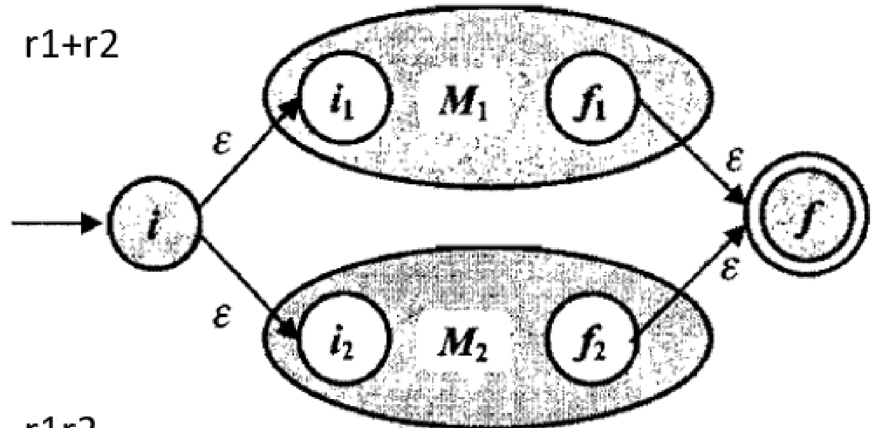
ε



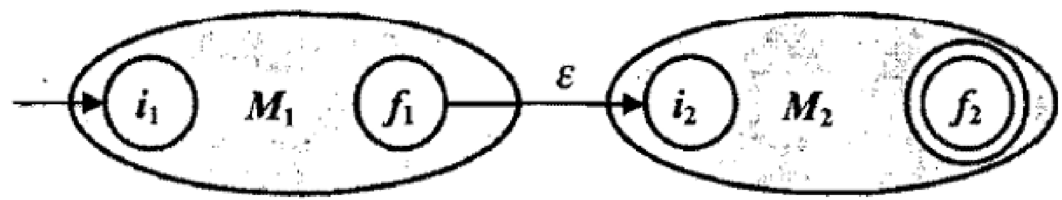
a



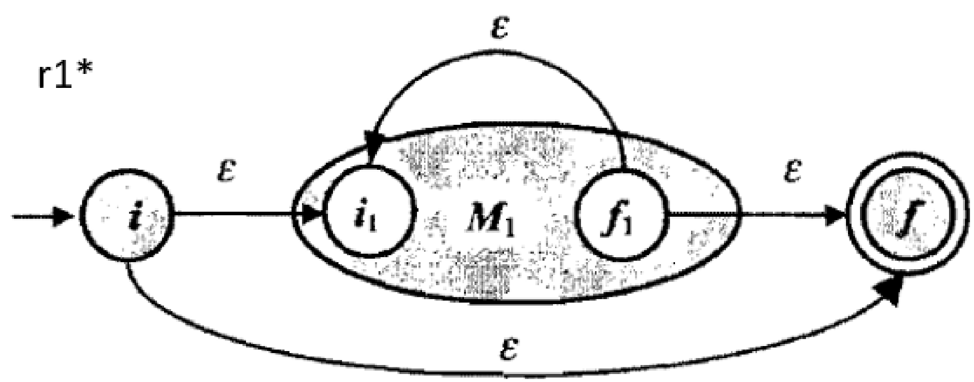
r1+r2



r1r2



r1*

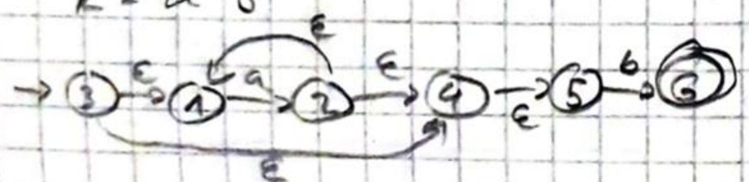


3

$R = ab$

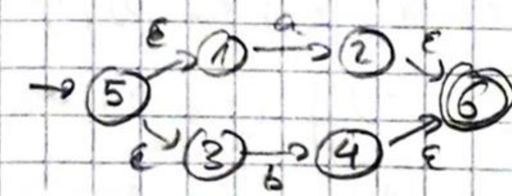
4

$R = a^*b$



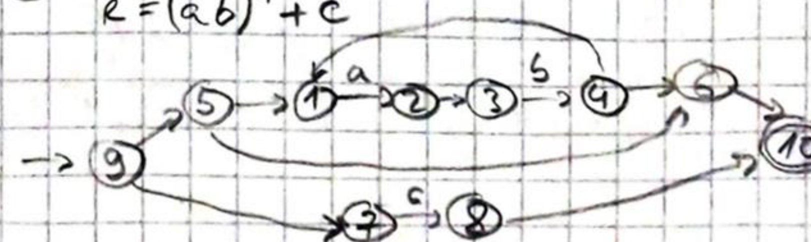
5

$R = a + b$



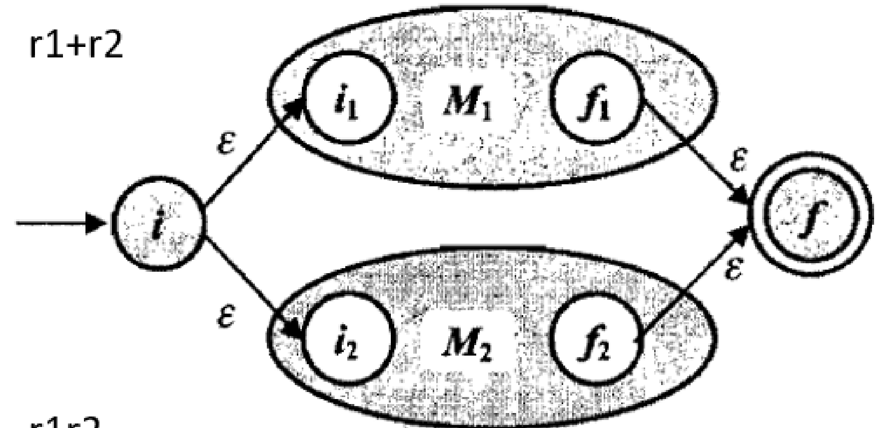
6

$R = (ab)^* + c$



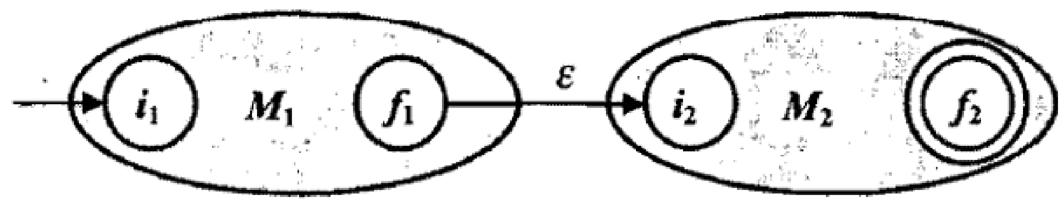
(7) R = a*b*   (8) R = (a+b)*+a*c   (9) R = ab+(a*+b*)*

r1+r2

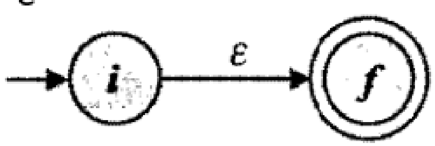


∅

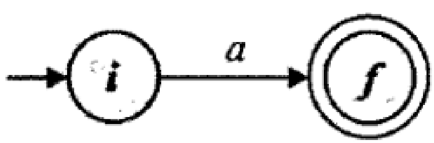
r1r2



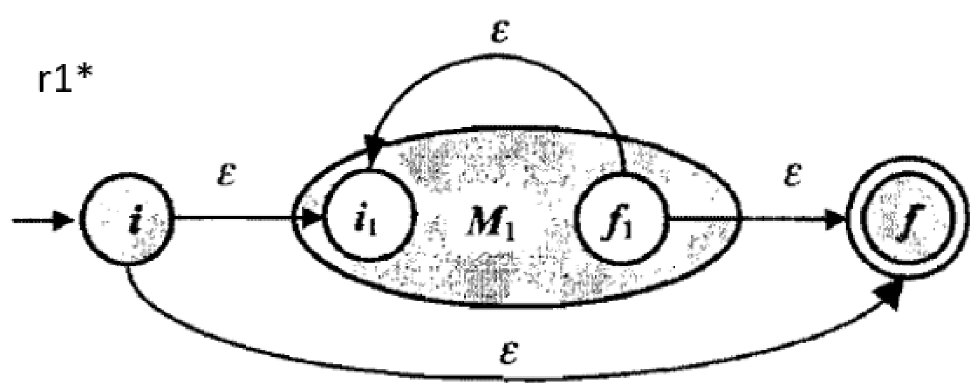
ε



a

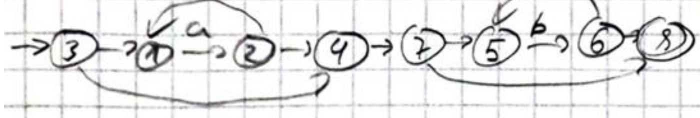


r1*



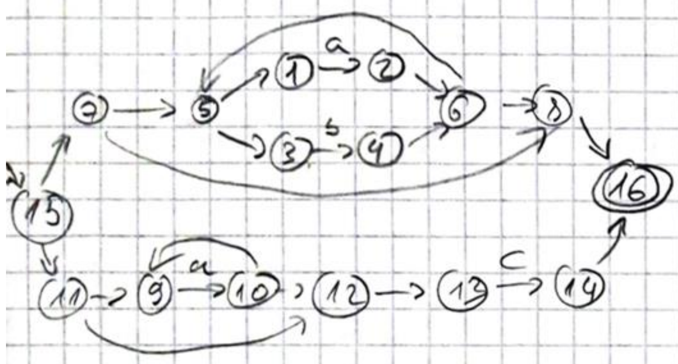
⑦

$R = a^*b^*$



⑧

$R = (a+b)^* + a^*c$



⑨

$R = ab + (a^* + b^*)^*$