Tavlenoter
Forelæsning om regulære udtryk

Christian Jødal O'Keeffe

14. februar 2013

Indhold

1 Indledning **1**

2 Regulære udtryk **2**
2.1 Definition af regulære udtryk 2
2.2 Forkortelser af regulære udtryk 2
2.3 Eksempler på regulære udtryk 2

3 Fra regulære udtryk til NFA **2**
3.1 Eksempel 3

4 Fra DFA til regulære udtryk **10**

5 Gyser **15**
5.1 Termolig 15
5.2 Ingen ad-hoc-løsninger, tak! 15
5.3 Ingen smarte genveje, tak! 15

1 Indledning

Denne forelæsning handler om

- Definition af regulære udtryk
- Eksempler på regulære udtryk
- Kleene's sætning (del 1)
  - Fra regulære udtryk til NFA
  - Fra DFA til regulære udtryk
- Gyser

2 Regulære udtryk

2.1 Definition af regulære udtryk

Definition 1 (Regulære udtryk)**.** Givet alfabetet Σ er mængden af regulære udtryk over Σ givet ved:

Regulært udtryk R	**Sproget beregnet af det regulære udtryk** $L(R)$
$\mathbf{a}$ (for $a \in \Sigma$)	$\{a\}$
$\emptyset$	$\{\}$
ϵ	$\{\epsilon\}$
$(R_1 \cup R_2)$	$L(R_1) \cup L(R_2)$
$(R_1 \circ R_2)$	$L(R_1) \circ L(R_2)$
$(R)^*$	$(L(R))^*$

2.2 Forkortelser af regulære udtryk

Man forkorter mange gange regulære udtryk når de skrives op. Ved forkortelser udlades $\circ$ tit samt diverse parenteser. Som eksempel kan det regulære udtryk

$$(\mathbf{1})^* \circ \mathbf{0}$$

skrives som

$$\mathbf{1}^*\mathbf{0}$$

Den forkortede notation vil blive brugt i de videre noter.

2.3 Eksempler på regulære udtryk

Lad $\Sigma = \{0, 1\}$

Så vil:

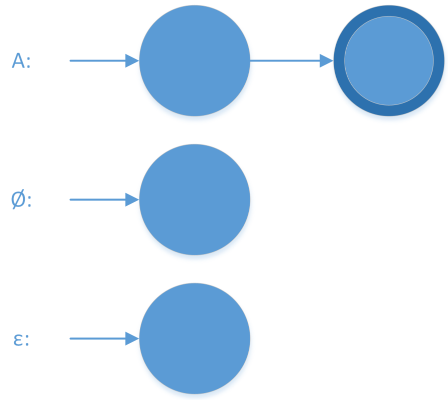
- $\mathbf{1}^*\mathbf{0}(\mathbf{1}^*\mathbf{0}\mathbf{1}^*\mathbf{0}\mathbf{1}^*)^* \cup (\mathbf{1} \cup \mathbf{0})^*\mathbf{11}$ beregne sproget af alle de strenge der har et ulige antal 0'er eller slutter på 11.
- $\mathbf{\Sigma\Sigma\Sigma}$ betegne alle strenge af længden 3 (kan også skrives $(1\cup 0)(1\cup 0)(1\cup 0)$).
- $\mathbf{0\Sigma}^*$ betegne alle strenge der begynder med 0.

3 Fra regulære udtryk til NFA

Sætning 1. *For ethvert regulært udtryk R findes en NFA N så $L(N) = L(R)$.*

Bevis: Induktion i længden af R, k

Basis: Ved $k = 1$ vil der være 3 tilfælde set på Figur 1.



Figur 1: Tre tilfælde ved basistrinnet

Skridt – antag påstand for alle $k' < k$, vis for k: Der er 3 tilfælde, som alle gør brug af at de regulære sprog er lukket under de regulære operationer.

$R = (R_1 \cup R_2)$: Da der findes per induktionsantagelse en N_1, N_2 så $L(R_1) = L(N_1)$, $L(R_2) = L(N_2)$. Vi skal lave en NFA der genkender $L(N_1) \cup L(N_2)$, og det er muligt da de regulære udtryk er lukket under $\cup$.

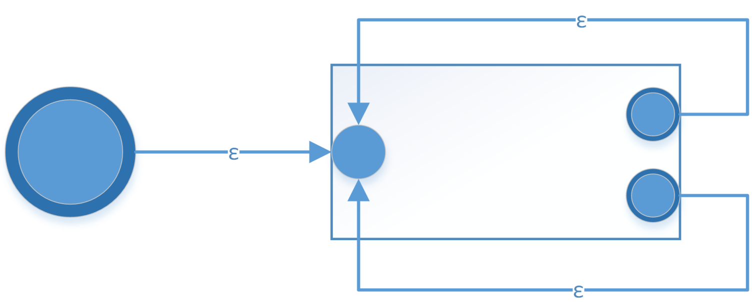
$R = (R_1 \circ R_2)$: Igen findes pr. induktionsantagelse N_1, N_2 så $L(R_1) = L(N_1), L(R_2) = L(N_2)$, og det er muligt at konstruere en NFA N så $L(N_1) \circ L(N_2)$. de regulære sprog er lukket under $\circ$.

$(R)^*$: Pr. induktionsantagelse findes en NFA N så $L(R) = L(N)$. Vi kan lave en NFA som genkender $(L(N))^*$ da de regulære udtryk er lukket under $*$

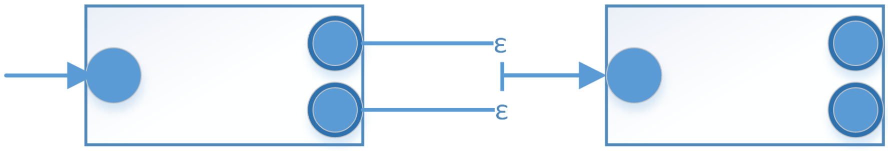
□

3.1 Eksempel

Husk fra tidligere forelæsning konstruktionen af en NFA for L^* (se Figur 2) samt $\circ$-konstruktionen (se Figur 3).



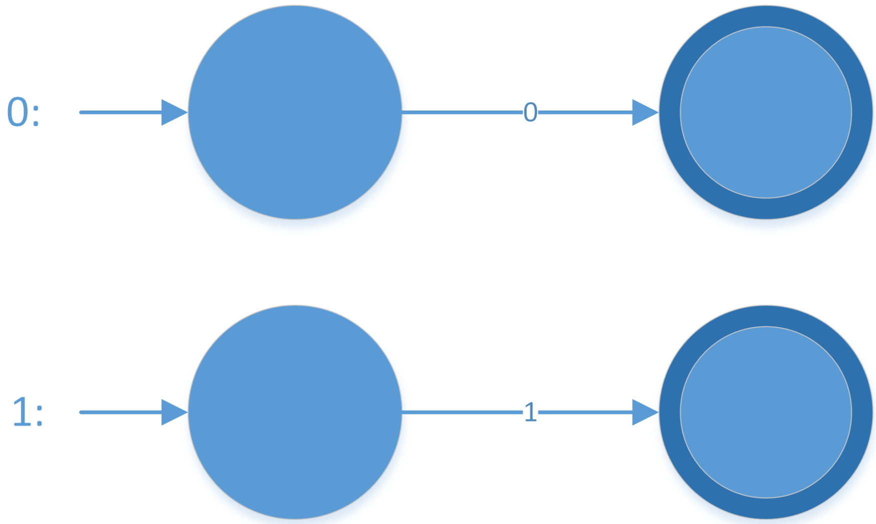
Figur 2: *-konstruktion af NFA



Figur 3: $\circ$-konstruktionen af NFA

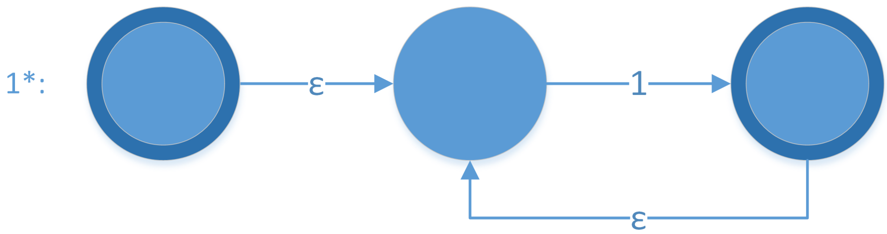
Eksempel 1. Betragt det regulære udtryk $\mathbf{1}^*\mathbf{0}(\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*)^* \cup (\mathbf{1} \cup \mathbf{0})^*\mathbf{11}$

For at konstruere en NFA fra dette regulære udtryk starter man med at konstruere NFA'er for de mindste bestanddele, og anvender derefter konstruktionerne for de regulære operationer. I dette tilfælde starter man altså med at konstruere en NFA for **0** og **1**, som set på Figur 4.



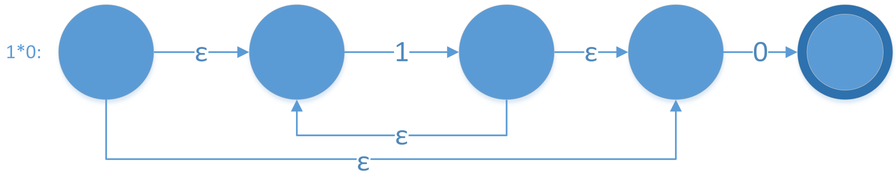
Figur 4: NFA'er for **0** og **1**

Herefter konstrueres NFA'en for $\mathbf{1}^*$ som set på Figur 5.



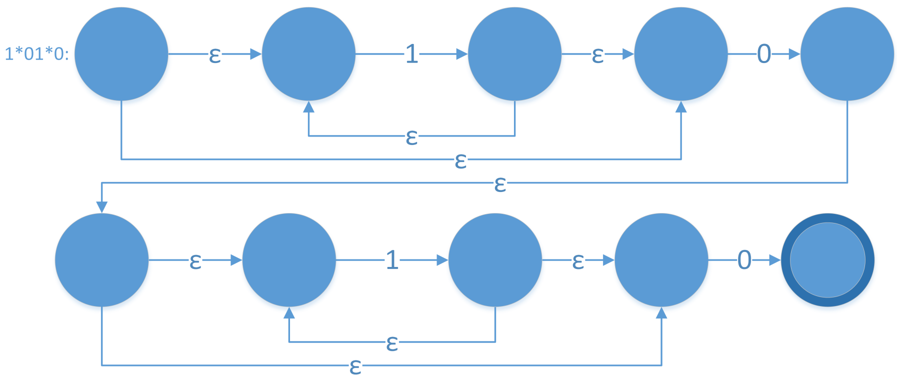
Figur 5: NFA for $\mathbf{1}^*$

Herefter konstrueres NFA'en for **1***0** som set på Figur 6.



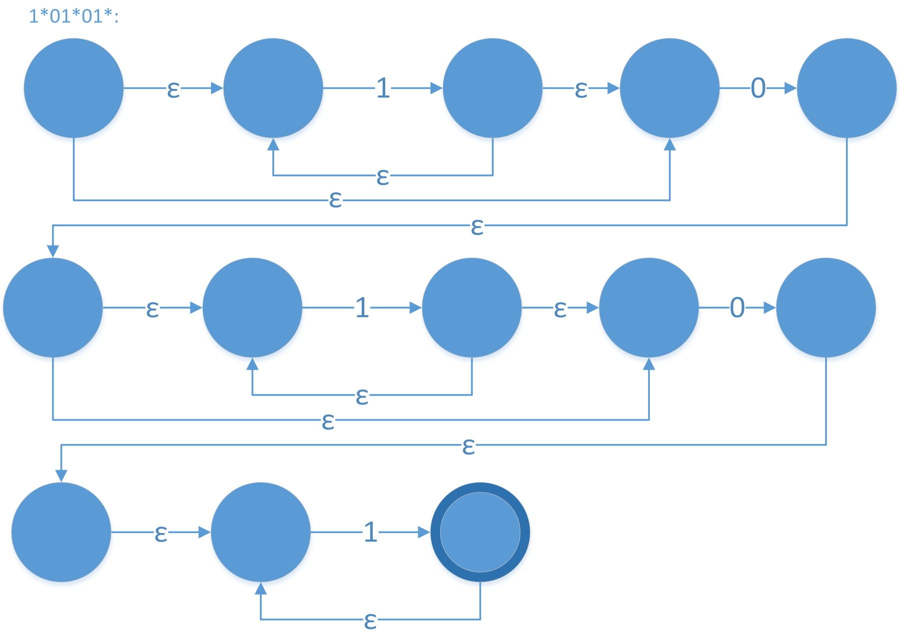
Figur 6: NFA for **1***0**

Herefter konstrueres NFA'en for **1***01***0** som set på Figur 7.



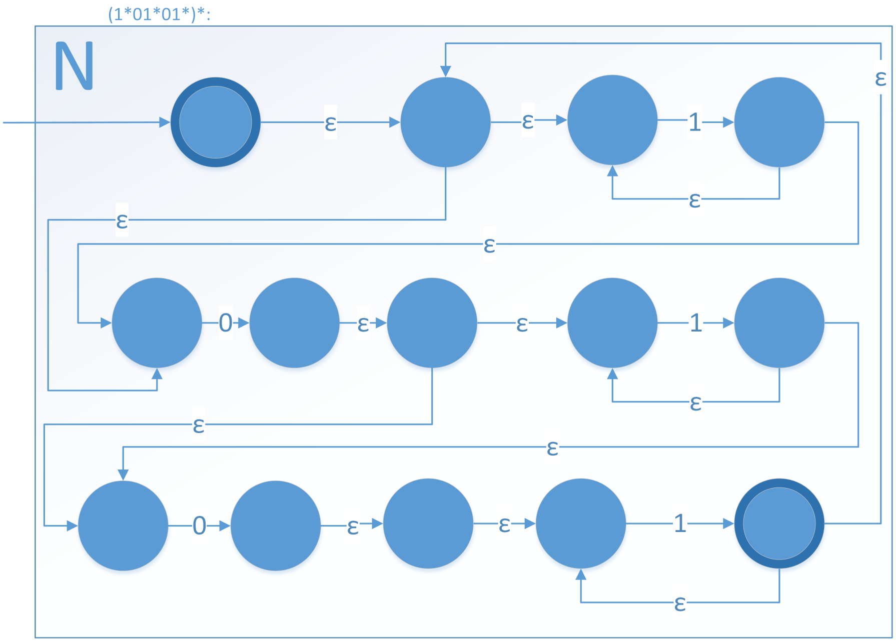
Figur 7: NFA for **1***01***0**

Herefter konstrueres NFA'en for **1***01***01*** som set på Figur 8.



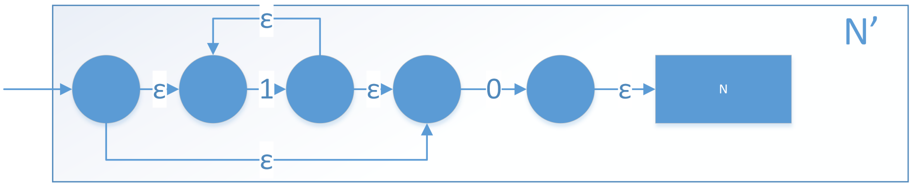
Figur 8: NFA for $\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*$

Herefter konstrueres NFA'en for $(\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*)^*$ som set på Figur 9. Denne NFA kaldes fremover N.



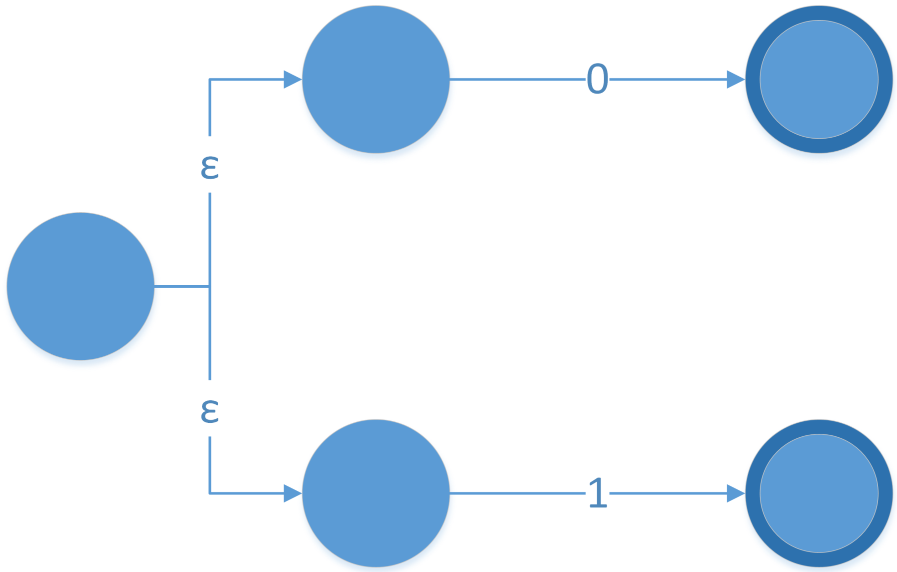
Figur 9: NFA for $(\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*)^*$

Herefter konstrueres NFA'en for $\mathbf{1} * \mathbf{0}(\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*)^*$ som set på Figur 10. Denne NFA kaldes fremover N'.



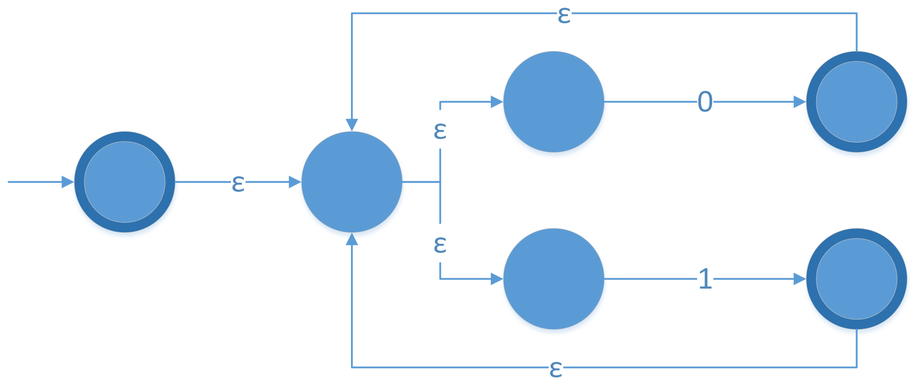
Figur 10: NFA for $\mathbf{1}^*\mathbf{0}(\mathbf{1}^*\mathbf{01}^*\mathbf{01}^*)^*$

Herefter konstrueres NFA'en for $\mathbf{0} \cup \mathbf{1}$ som set på Figur 11.



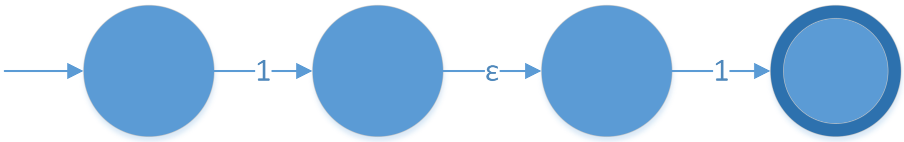
Figur 11: NFA for $\mathbf{0} \cup \mathbf{1}$

Herefter konstrueres NFA'en for $(\mathbf{0} \cup \mathbf{1})^*$ som set på Figur 12.



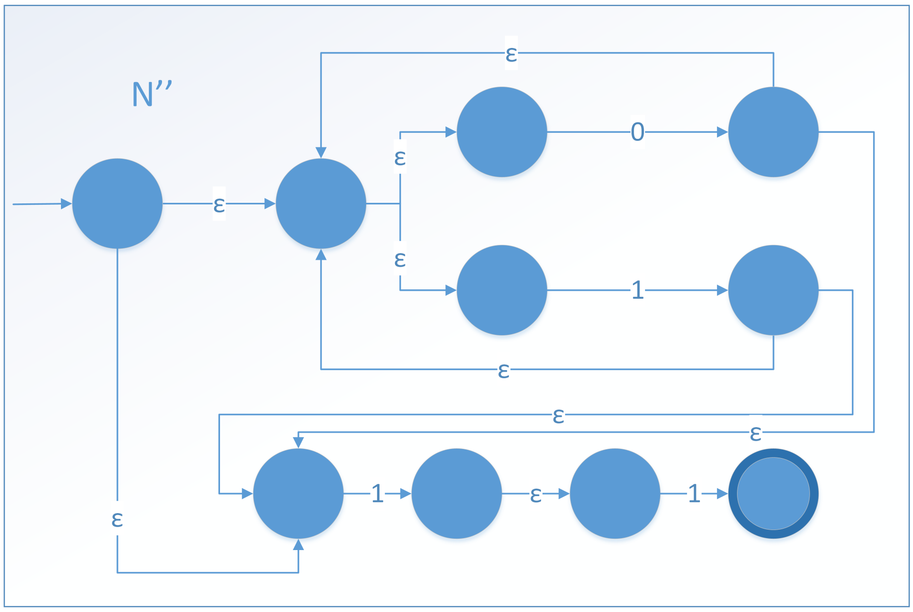
Figur 12: NFA for $(\mathbf{0} \cup \mathbf{1})^*$

Herefter konstrueres NFA'en for $\mathbf{11}$ som set på Figur 13.



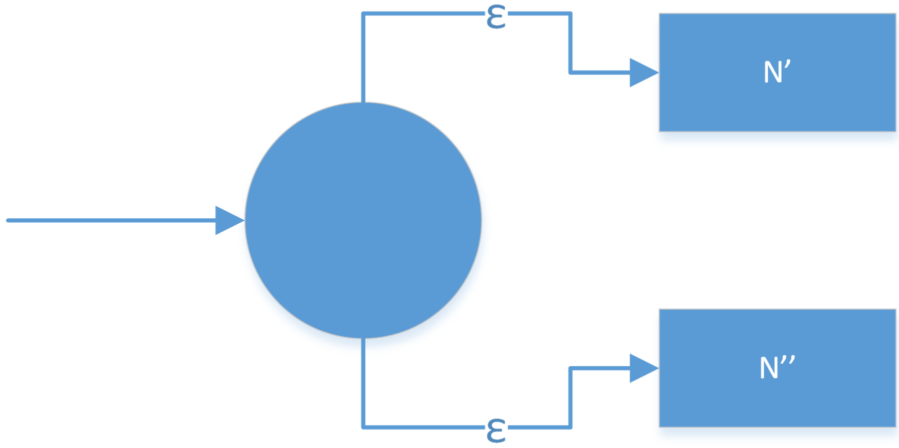
Figur 13: NFA for $\mathbf{11}$

Herefter konstrueres NFA'en for $(\mathbf{0} \cup \mathbf{1})^*\mathbf{11}$ som set på Figur 14. Denne NFA kaldes fremover N''.



Figur 14: NFA for $(\mathbf{0} \cup \mathbf{1})^*\mathbf{11}$

Til slut sammensættes N' og N'' som set på Figur 15.



Figur 15: NFA for $\mathbf{1}^*\mathbf{0}(\mathbf{1}^*\mathbf{0}\mathbf{1}^*\mathbf{0}\mathbf{1}^*)^* \cup (\mathbf{1} \cup \mathbf{0})^*\mathbf{11}$

4 Fra DFA til regulære udtryk

Sætning 2. *For enhver DFA M findes der et regulært udtryk R så $L(R) = L(M)$.*

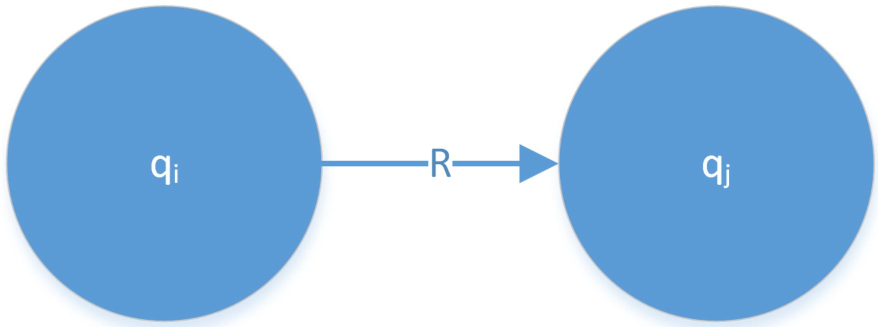
Bevisideen er denne:

For M:

- Lav M om til en automat med regulære udtryk på transitionerne, G_M.
- Fjern tilstande fra G_M en ad gangen (men sådan at samme sprog genkendes).
- Til sidst er der en en start og en sluttilstand, med et regulært udtryk på transitionen fra start til slut.

En automat med regulære udtryk på transitionerne kaldes en *generaliseret NFA* (GNFA).

Definition 2. En GNFA er en 5-tupel $(Q, \Sigma, q_{start}, q_{accept}, \delta)$ hvor der står regulære udtryk på transitionerne (se Figur 16).



Figur 16: En GNFA med regulære udtryk på transitionerne

Overføringsfunktion fortæller for hvert par af tilstande hvilket regulært udtryk der er på transitionen mellem dem, dvs.

$$\delta : Q \setminus \{q_{accept}\} \times Q \setminus \{q_{start}\} \to \mathcal{R}$$

hvor $\mathcal{R}$ betegner mængden af regulære udtryk.

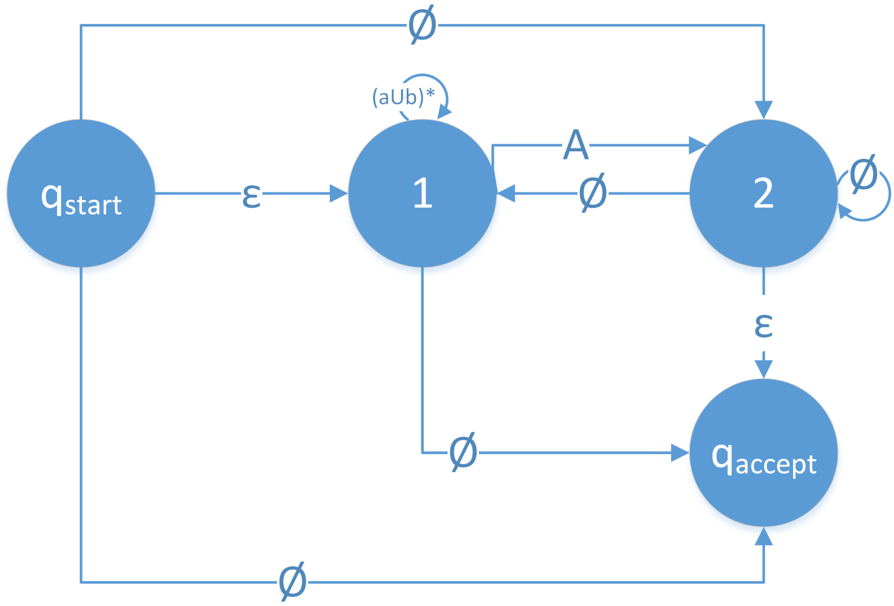
En GNFA overholder denne bordskik:

1. En transition mellem hvert par af tilstande, dog skal det gælde at
2. Der er ingen transitioner *fra* q_{accept}
3. Der er ingen transitioner *til* q_{start}

En GNFA accepterer en streng w hvis $w = s_1 \ldots s_n$ og der er en følge af tilstande $r_1 \ldots r_n$ så:

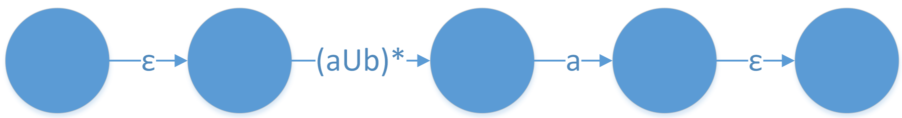
- $r_1 = q_{start}$
- For alle $r_i (1 \leq i < n)$ gælder $\delta(r_i, r_{i+1}) = R_i$ og $s_i \in L(R_i)$
- $r_n = q_{accept}$

Eksempel 2. På Figur 17 ses et eksempel på en GNFA.



Figur 17: Et eksempel på en GNFA

Her accepteres fx strengen *aaba*. Dette kan vises ved at *aaba* kan deles op i ϵ *aab* *a* ϵ, se Figur 18.



Figur 18: Sådan kan strengen *aaba* læses af GNFA'en fra Figur 17

Bevis: (for Sætning 2) Givet en DFA M der ønsket skrevet i form at regulære udtryk. Først skal M laves om til en GNFA. Dette gøres ved:

- Indfør ny q_{start} og en transition ϵ til det gamle startpunkt ved M
- Indfør ny q_{accept} og lav for hver gammen accepttilstand en ϵ transition til q_{accept}
- Hvis der ikke er en transition mellem q_i og q_j i M tilføjes en transition mærket $\emptyset$
- Saml transitioner med $\cup$ så en transition mærket a, b i stedet bliver mærket $a \cup b$

Herefter fjernes tilstandene en efter en, så der kun er q_{start} og q_{accept} tilbage. Hver gang en tilstand fjernes fra M og vi får en ny automat M', skal vi sikre at $L(M') = L(M)$, dvs. at vores nye automat genkender samme sprog som den gamle.

Dette gør vi således. Lad q_{rip} være den tilstand som vi fjerner. Det skal gælde at $q_{rip} \neq q_{start}$ og $q_{rip} \neq q_{accept}$.

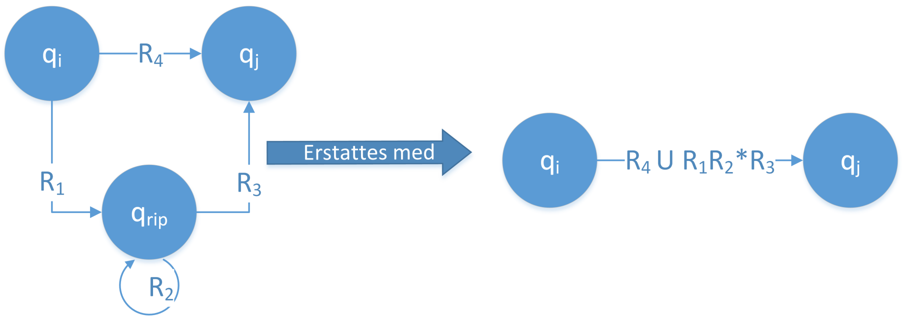
Da vores GNFA M overholder bordskik, er der en transition mellem hvert par af tilstande. Betragt to vilkårlige tilstande q_i, q_j som også skal findes i M'. På grund af vores krav om bordskik ved vi at der er

- en transition fra q_i til q_j mærket med R_4
- en transition fra q_i til q_{rip} mærket med R_1
- en transition fra q_{rip} til q_{rip} mærket med R_2
- en transition fra q_{rip} til q_j mærket med R_3

Hvad for et regulært udtryk skal der stå på transitionen i M' fra q_i til q_J? Der skal stå et regulært udtryk der beskriver alle de strenge, der kan føre os fra q_I til q_j. Dette regulære udtryk er

$$R_4 \cup R_1(R_2)^* R_3$$

Dette kan ses på Figur 19. Bemærk at dette skal gøres *for hvert par af tilstande* q_i, q_j, herunder i tilfældene hvor $q_i = q_j$ (der skal være en tilstand mellem hvert par af tilstande, herunder fra hver tilstand, der ikke er q_{start} eller q_{accept}, til sig selv).



Figur 19: Sådan laves den nye transition mellem q_i og q_j

Den samlede konverteringsalgoritme er

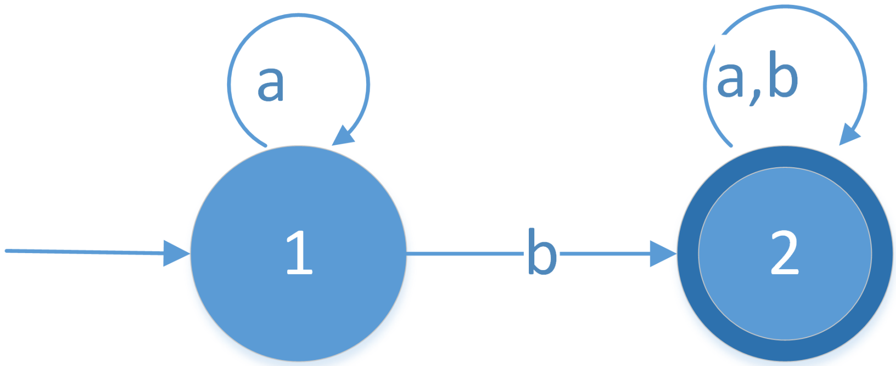
CONVERT(G) =

1. Lad Q_G betegne mængden af tilstande i G. Hvis de eneste tilstande i Q_G er q_{start} og q_{accept}, så returnér det regulære udtryk på transitionen mellem q_{start} og q_{accept}.
2. Ellers vælg en tilstand $q_{rip} \in Q_G$ hvor $q_{rip} \neq q_{start}$ og $q_{rip} \neq q_{accept}$. Lad mængden af tilstande i den nye automat være $Q_G \setminus \{q_{rip}\}$.

3. For hvert par af tilstande q_i, q_j skab den nye transition som angivet på Figur 19.
4. Kald den nye automat for G'.
5. Kald $CONVERT(G')$.

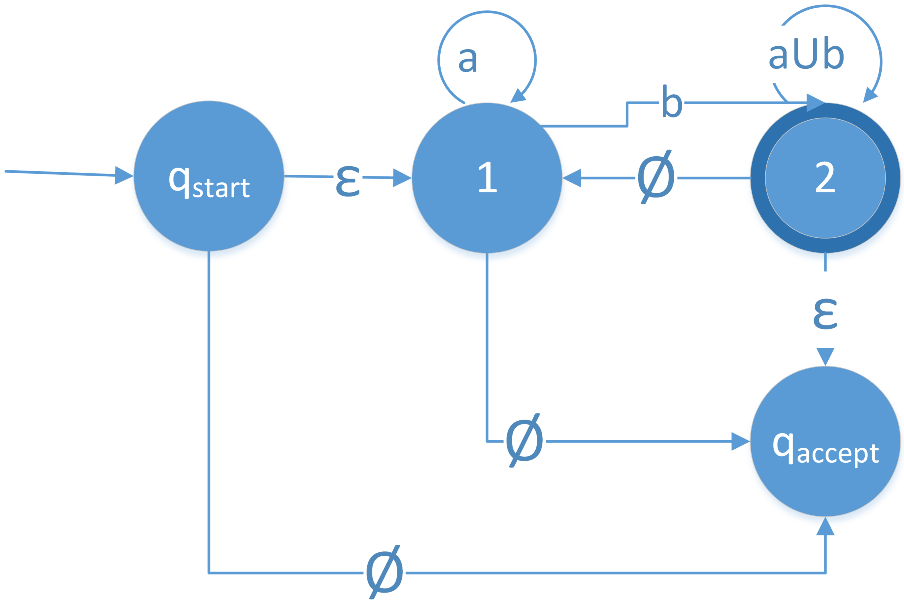
□

Eksempel 3. Givet er en DFA kaldet M, som kan ses på Figur 20.



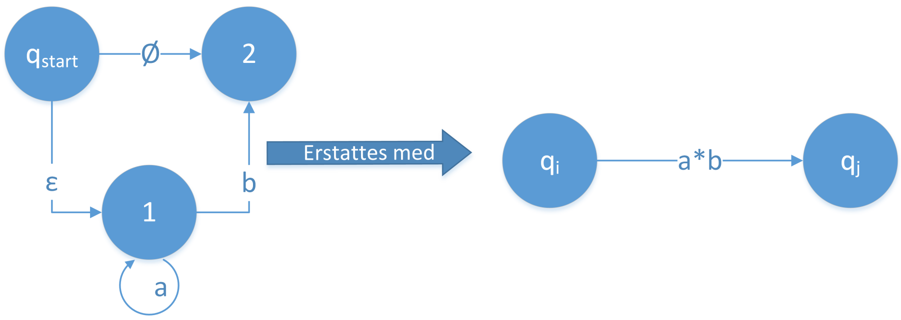
Figur 20: En DFA M

Herefter laves en GNFA ud fra DFA'en M, som kan ses på Figur 21.



Figur 21: En GNFA ud fra DFA'en M

Først vælges tilstand 1 som q_{rip} som ses på 22.



Figur 22: Udledningen af det regulære udtryk der skal erstatte tilstand 1

Dernæst vælges tilstand 2 som q_{rip} og resultatet er en GNFA hvis eneste transition er mærket med det regulære udtryk $a^*b(a \cup b)^*$.

5 Gyser

5.1 Termolig

Forbløffende mange har problemer med terminologien også her.

- Regulære udtryk *beskriver* sprog, automater *genkender* sprog.
  - ikke "regulære udtryk genkender"
  - ikke "regulære udtryk accepterer"

5.2 Ingen ad-hoc-løsninger, tak!

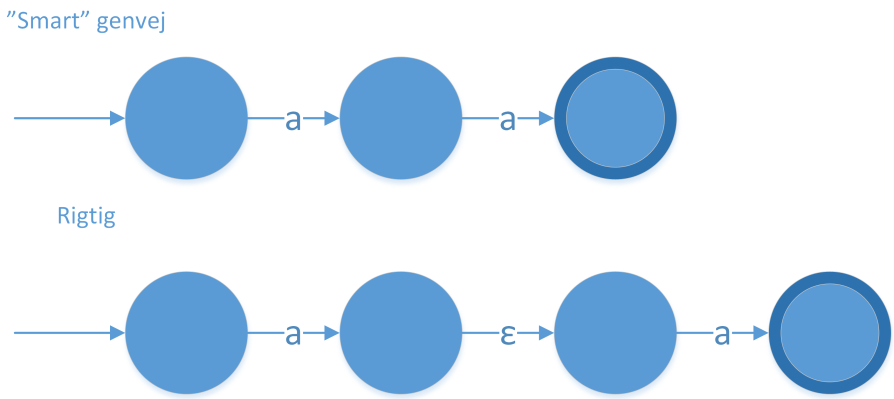
Nogle studerende vil af en eller anden grund ikke bruge tid på at lære de generelle konstruktioner som dukker op i kurset. Når de til eksamen bliver bedt om at konstruere en NFA ud fra et regulært udtryk eller bliver bedt om at et regulært udtryk ud fra en DFA, prøver de at finde automaten eller det regulære udtryk ved at gætte sig frem.

Men formålet med denne type opgave er *at finde ud af om den studerende kender den generelle konstruktion*. Noget helt andet er at de studerende, der ad hoc'er sig frem, aldrig får fundet den rigtige automat eller det rigtige regulære udtryk!

5.3 Ingen smarte genveje, tak!

Nogle studerende tror at man kan "spare" ved at lade være med at bruge de korrekte konstruktioner og gøre som f.eks. på Figur 23.

Men den slags fører meget, meget let til fejl – og man glemmer meget, meget let hvad den generelle metode består i. Gør det rigtigt fra starten, og undgå "smarte genveje" der ignorerer den generelle metode.



Figur 23: Undgå smarte genveje som denne.