

# Syntaks og semantik

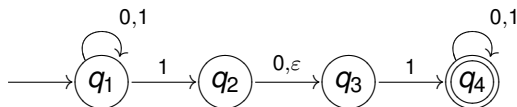
## Lektion 4

20 februar 2007

# Forord

- 1 Administrivia
- 2 Non-deterministiske endelige automater
- 3 NFAs og regulære udtryk

- Der skulle nu være nok *Sipsere* i boghandelen
- Deadline for aflevering af syntaksopgave-erstatnings-opgavestilling (for PE-studerende) er *i dag*!
- næste gang: spørgetime!



**Definition 1.37:** En **nondeterministisk endelig automat (NFA)** er en 5-tupel  $M = (Q, \Sigma, \delta, q_0, F)$ , hvor delene er

- 1  $Q$  : en endelig mængde af tilstande
- 2  $\Sigma$  : input-alfabetet
- 3  $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow \mathcal{P}(Q)$  : transitions-funktionen
- 4  $q_0 \in Q$  : starttilstanden
- 5  $F \subseteq Q$  : mængden af accepttilstande

$M$  siges at **acceptere** et ord  $w \in \Sigma^*$  hvis der findes  $m \in \mathbb{N}$  og  $y_1, y_2, \dots, y_m \in \Sigma \cup \{\varepsilon\}$  og  $r_0, r_1, \dots, r_m \in Q$  således at  $w = y_1 y_2 \dots y_m$  og

- 1  $r_0 = q_0$ ,
- 2  $r_{i+1} \in \delta(r_i, y_{i+1})$  for alle  $i = 0, 1, \dots, m-1$ , og
- 3  $r_m \in F$ .

- enhver DFA er også en NFA
  - enhver NFA kan laves om til en DFA der genkender samme sprog (**delmængdekonstruktionen**)
  - et sprog er defineret til at være **regulært** hvis der er en DFA der genkender det
- ⇒ et sprog er regulært hvis og kun hvis der er en NFA der genkender det
- regulære sprog er lukket under  $\cup$ ,  $\circ$ ,  $*$  (vises ved at konstruere en ny NFA ud fra de givne NFAs)
  - regulære sprog er lukket under  $\cap$  og  $^-$  (komplement) (vises ved at konstruere en ny **DFA** ud fra de givne **DFAs**; konstruktionerne virker *kun* for DFAs!)
  - NFAs er generelt mere simple at fremstille
  - men nogen gange kan det være nødvendigt at arbejde med DFAs – eksempel: opgave 1.13

**Lemma 1.55:** Hvis et sprog beskrives ved et regulært udtryk, da er det regulært.

**Bevises ved strukturel induktion:**

- 1 konvertér de basale regulære udtryk til NFAs
- 2 brug lukningsegenskaber til at konvertere sammensætninger af regulære udtryk til sammensætninger af NFAs
- 3 **Smart!**

I dag: **Lemma 1.60:** Hvis et sprog er regulært, da kan det beskrives ved et regulært udtryk.

(Bevises ved at *generalisere* NFAs til **GNFAs**.)

⇒ **Sætning 1.54:** Et sprog er regulært **hvis og kun hvis** det kan beskrives ved et regulært udtryk.

# Regulære og ikke-regulære sprog

- 4 Regulære sprog genereres af regulære udtryk
- 5 Ikke-regulære sprog

**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Nøgle** til beviset: Ny slags maskiner der kombinerer NFA og regulære udtryk: **generaliserede nondeterministiske endelige automater (GNFA)**

**Definition 1.64:** En **GNFA** er en 5-tupel  $(Q, \Sigma, \delta, q_0, q_f)$ , hvor delene er

- 1  $Q$  : en endelig mængde af tilstande
- 2  $\Sigma$  : input-alfabetet
- 3  $\delta : (Q \setminus \{q_f\}) \times (Q \setminus \{q_0\}) \rightarrow \mathcal{R}$  : transitions-funktionen
- 4  $q_0 \in Q$  : starttilstanden
- 5  $q_f \in Q$  : accepttilstanden

**Notation:**  $\mathcal{R} = \mathcal{R}(\Sigma)$  = mængden af alle **regulære udtryk** over et givet alfabet  $\Sigma$ .

(Bemærk at GNFA's introduceres kun for det her bevis. De bruges ikke til andet.)

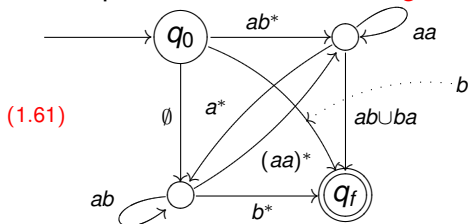


Definition 1.64: En **GNFA** er en 5-tupel  $(Q, \Sigma, \delta, q_0, q_f)$ , hvor delene er

- ③  $\delta : (Q \setminus \{q_f\}) \times (Q \setminus \{q_0\}) \rightarrow \mathcal{R} : \text{transitions-funktionen}$
- ⑤  $q_f \in Q : \text{accepttilstanden}$

Ligesom NFAs, men

- med kun **én** accepttilstand
- med **regulære udtryk** på transitionerne i stedet for tegn
- med transitioner **fra enhver tilstand til enhver tilstand** (også sig selv), bortset fra at
  - starttilstanden ikke har **indgående** transitioner, og at
  - accepttilstanden ikke har **udgående** transitioner



Definition 1.64: En **GNFA** er en 5-tupel  $(Q, \Sigma, \delta, q_0, q_f)$ , hvor delene er

- 1  $Q$  : en endelig mængde af tilstande
- 2  $\Sigma$  : input-alfabetet
- 3  $\delta : (Q \setminus \{q_f\}) \times (Q \setminus \{q_0\}) \rightarrow \mathcal{R}$  : transitions-funktionen
- 4  $q_0 \in Q$  : starttilstanden
- 5  $q_f \in Q$  : accepttilstanden

GNFAen **accepterer** et ord  $w \in \Sigma^*$  hvis der findes  $m \in \mathbb{N}$  og  $y_1, y_2, \dots, y_m \in \Sigma^*$  (!) og  $r_0, r_1, \dots, r_m \in Q$  således at  $w = y_1 y_2 \dots y_m$  og

- 1  $r_0 = q_0$ ,
- 2  $y_{i+1} \in \llbracket \delta(r_i, r_{i+1}) \rrbracket$  for alle  $i = 0, 1, \dots, m-1$ , og
- 3  $r_m = q_f$ .

**Bevisidé:** konvertér en DFA til en GNFA og så GNFAen til et regulært udtryk ved at **fjerne én tilstand ad gangen**.

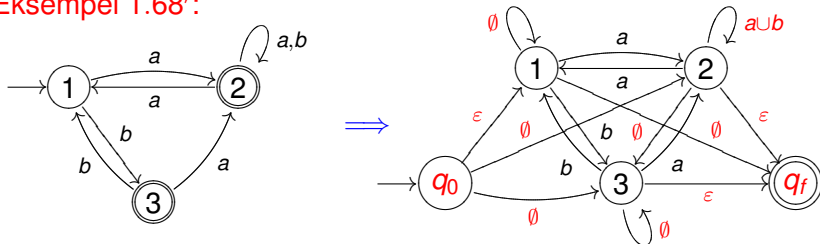
**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Bevis:** Lad  $M = (Q_1, \Sigma, \delta_1, q_1, F_1)$  være en DFA med  $\llbracket M \rrbracket = L$ .

① Konvertér  $M$  til en GNFA  $G = (Q, \Sigma, \delta, q_0, q_f)$ :

- (a) Lav en ny starttilstand  $q_0$  og en ny accepttilstand  $q_f$ , med  $\varepsilon$ -transitioner fra  $q_0$  til den gamle starttilstand og fra alle gamle accepttilstande til  $q_f$ .
- (b) Erstat transitioner med flere end ét label med én transition der som label har foreningen af disse labels.
- (c) Indsæt  $\emptyset$ -transitioner hvor der mangler pile.

**Eksempel 1.68':**



**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Bevis:** Lad  $M = (Q_1, \Sigma, \delta_1, q_1, F_1)$  være en DFA med  $\llbracket M \rrbracket = L$ .

① Konvertér  $M$  til en GNFA  $G = (Q, \Sigma, \delta, q_0, q_f)$ :

- (a) Lav en ny starttilstand  $q_0$  og en ny accepttilstand  $q_f$ , med  $\varepsilon$ -transitioner fra  $q_0$  til den gamle starttilstand og fra alle gamle accepttilstande til  $q_f$ .
- (b) Erstat transitioner med flere end ét label med én transition der som label har foreningen af disse labels.
- (c) Indsæt  $\emptyset$ -transitioner hvor der mangler pile.

$$Q = Q_1 \cup \{q_0, q_f\}$$

$$\delta(q, q') = \begin{cases} \varepsilon & \text{hvis } q = q_0 \text{ eller } q' = q_f & (a) \\ a_1 \cup a_2 \cup \dots \cup a_k & \text{hvis } q, q' \in Q_1 \text{ og } \delta_1(q, a_i) = q' \\ & \text{for alle } i = 1, 2, \dots, k & (b) \\ \emptyset & \text{hvis } q, q' \in Q_1 \text{ og } \delta_1(q, a) \neq q' \\ & \text{for alle } a \in \Sigma & (c) \end{cases}$$

**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Bevis:** Lad  $M = (Q, \Sigma, \delta, q_0, F)$  være en DFA med  $\llbracket M \rrbracket = L$ .

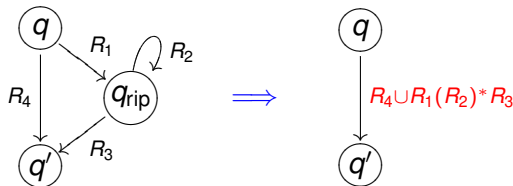
- 1 Konvertér  $M$  til en GNFA  $G = (Q, \Sigma, \delta, q_0, q_f)$
- 2 Konvertér  $G$  til et regulært udtryk  $R$ :

**CONVERT**( $G$ ):

- 1 Lad  $k = |Q|$  – antallet af tilstande i  $G$ .
- 2 Hvis  $k = 2$ , returnér  $\delta(q_0, q_f)$ .
- 3 Vi har  $k > 2$ . Lad  $q_{\text{rip}} \in Q \setminus \{q_0, q_f\}$ .

Lad  $Q' = Q \setminus \{q_{\text{rip}}\}$ , og definér

$\delta' : (Q' \setminus \{q_f\}) \times (Q' \setminus \{q_0\}) \rightarrow \mathcal{R}$  på følgende måde:



**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Bevis:** Lad  $M = (Q, \Sigma, \delta, q_0, q_f)$  være en DFA med  $\llbracket M \rrbracket = L$ .

- 1 Konvertér  $M$  til en GNFA  $G = (Q, \Sigma, \delta, q_0, q_f)$
- 2 Konvertér  $G$  til et regulært udtryk  $R$ :

**CONVERT**( $G$ ):

- 1 Lad  $k = |Q|$  – antallet af tilstande i  $G$ .
- 2 Hvis  $k = 2$ , returnér  $\delta(q_0, q_f)$ .
- 3 Vi har  $k > 2$ . Lad  $q_{\text{rip}} \in Q \setminus \{q_0, q_f\}$ .

Lad  $Q' = Q \setminus \{q_{\text{rip}}\}$ , og definér

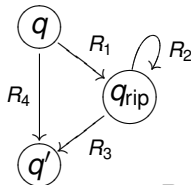
$\delta' : (Q' \setminus \{q_f\}) \times (Q' \setminus \{q_0\}) \rightarrow \mathcal{R}$  på følgende måde:

For  $q \in Q' \setminus \{q_f\}$  og  $q' \in Q' \setminus \{q_0\}$  lad

$R_1 = \delta(q, q_{\text{rip}})$ ,  $R_2 = \delta(q_{\text{rip}}, q_{\text{rip}})$ ,

$R_3 = \delta(q_{\text{rip}}, q')$  og  $R_4 = \delta(q, q')$ , og lad

$\delta'(q, q') = R_4 \cup R_1(R_2)^*R_3$ .



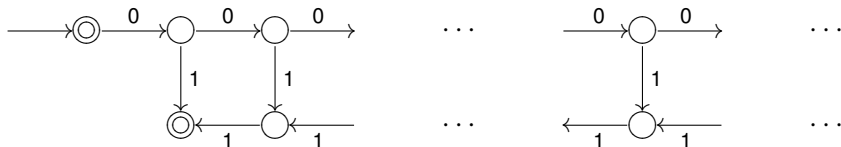
- 4 Returnér **CONVERT**( $G' = (Q', \Sigma, \delta', q_0, q_f)$ )

**Lemma 1.60:** Givet et alfabet  $\Sigma$  og et regulært sprog  $L \subseteq \Sigma^*$ , da findes et regulært udtryk  $R$  over  $\Sigma$  således at  $L = \llbracket R \rrbracket$ .

**Bevis:** Lad  $M = (Q_1, \Sigma, \delta_1, q_1, F_1)$  være en DFA med  $\llbracket M \rrbracket = L$ .

- ① Konvertér  $M$  til en GNFA  $G = (Q, \Sigma, \delta, q_0, q_f)$
- ② Konvertér  $G$  til et regulært udtryk  $R$ .
- ③ Vis at  $\llbracket M \rrbracket = \llbracket R \rrbracket$ :
  - ① Vis at  $\llbracket M \rrbracket = \llbracket G \rrbracket$ : nemt
  - ② Vis at  $\llbracket G \rrbracket = \llbracket R \rrbracket$ :
    - ① Hvis  $k = |Q| = 2$ :  $Q = \{q_0, q_f\}$ , og  $R = \delta(q_0, q_f)$   
 $\Rightarrow \checkmark$
    - ② Hvis  $k > 2$ : Vis at  $\llbracket G \rrbracket = \llbracket G' \rrbracket$
- ④ Done!

Ikke alle sprog er regulære. F.x. sproget  $\{0^n 1^n \mid n \in \mathbb{N}\}$ :



– en **uendelig** automat!

*Pumping Lemma*: en egenskab ved alle regulære sprog.

$\Rightarrow$  Hvis et sprog *ikke* har den egenskab, **kan det ikke være regulært**.



**Sætning 1.70 (Pumpelemmaet):** For ethvert regulært sprog  $A$  findes der et (naturligt) tal  $p$  således at ethvert ord  $s \in A$  der har længde mindst  $p$  kan opsplittes i tre stykker,  $s = xyz$ , med

- $|y| > 0$  og  $|xy| \leq p$ ,
- og således at ordene  $xy^i z \in A$  for alle  $i \in \mathbb{N}_0$ .

*En gang til:*

For ethvert regulært sprog  $A$

findes  $p \in \mathbb{N}_0$  således at

for ethvert  $s \in A$  med  $|s| \geq p$

findes en opsplitning  $s = xyz$  således at

$|y| > 0$  og  $|xy| \leq p$  og

for alle  $i \in \mathbb{N}_0$

$xy^i z \in A$ .

**Sætning 1.70 (Pumpelemmaet):** For ethvert regulært sprog  $A$  findes der et (naturligt) tal  $p$  således at ethvert ord  $s \in A$  der har længde mindst  $p$  kan opsplittes i tre stykker,  $s = xyz$ , med

- $|y| > 0$  og  $|xy| \leq p$ ,
- og således at ordene  $xy^iz \in A$  for alle  $i \in \mathbb{N}_0$ .

**Eksempel 1.73:** Sproget  $B = \{0^n1^n \mid n \in \mathbb{N}\}$  er ikke regulært.

**Bevis** (ved modstrid; kortere end i bogen!): *Antag* at  $B$  er regulært, og lad  $p$  være pumpelængden. Lad  $s = 0^p1^p$ , da er  $|s| \geq p$ .

Lad  $s = xyz$  være en opsplitning af  $s$  som opfylder pumpelemmaets betingelser. Pga.  $|xy| \leq p$  kan  $y$  kun indeholde 0'er, og pga.  $|y| > 0$  indeholder  $y$  mindst ét 0.

Sidste betingelse i lemmaet siger bl.a. at ordet  $xyyz \in A$ , men dette ord indeholder for mange 0'er. **Modstrid!**

**Sætning 1.70 (Pumpelemmaet):** For ethvert regulært sprog  $A$  findes der et (naturligt) tal  $p$  således at ethvert ord  $s \in A$  der har længde mindst  $p$  kan opsplittes i tre stykker,  $s = xyz$ , med

- $|y| > 0$  og  $|xy| \leq p$ ,
- og således at ordene  $xy^iz \in A$  for alle  $i \in \mathbb{N}_0$ .

**Bevis:** Lad  $M = (Q, \Sigma, \delta, q_0, F)$  være en DFA der genkender  $A$ , og lad  $p = |Q|$ . Lad  $s = s_1 s_2 \dots s_n \in A$  med  $|s| \geq p$ .

Mens  $M$  læser  $s$ , kommer den igennem en følge af  $n + 1$  tilstande. Men  $n + 1 > p$ , så der er flere tilstande i følgen end der er i  $M$ !

Dvs. der er en tilstand der optræder **to gange** i følgen – en **løkke**!

Hvis vi tager  $x$  til at være den del af  $s$  der læses *før* løkken,  $y$  den del der læses *i* løkken, og  $z$  den del der læses *efter* løkken, kan vi gennemløbe løkken  $i$  gange og genkende strengen  $xy^iz$ .

**Sætning 1.70 (Pumpelemmaet):** For ethvert regulært sprog  $A$  findes der et (naturligt) tal  $p$  således at ethvert ord  $s \in A$  der har længde mindst  $p$  kan opsplittes i tre stykker,  $s = xyz$ , med

- $|y| > 0$  og  $|xy| \leq p$ ,
- og således at ordene  $xy^iz \in A$  for alle  $i \in \mathbb{N}_0$ .

**Bevis:** Lad  $M = (Q, \Sigma, \delta, q_0, F)$  være en DFA der genkender  $A$ , og lad  $p = |Q|$ . Lad  $s = s_1 s_2 \dots s_n \in A$  med  $|s| \geq p$ .

Lad  $r_1, r_2, \dots, r_{n+1} \in Q$  således at  $r_1 = q_0$ ,  $r_{n+1} \in F$ , og  $r_{i+1} = \delta(r_i, s_i)$  for alle  $i$ .

Vi har  $n + 1 \geq p + 1$ , og  $|Q| = p$ . Derfor findes indices  $j$  og  $\ell$  således at  $1 \leq j < \ell \leq p + 1$  og  $r_j = r_\ell$ .

Lad  $x = s_1 \dots s_{j-1}$ ,  $y = s_j \dots s_{\ell-1}$ ,  $z = s_\ell \dots s_n$ . Pga.  $j < \ell$  har vi  $|y| \geq 0$ , og  $\ell \leq p + 1$  medfører  $|xy| \leq p$ .

Eftersom  $\delta(r_{\ell-1}, s_{\ell-1}) = r_j$ , er enhver følge

$(r_1, \dots, r_{j-1})(r_j, \dots, r_{\ell-1})^i(r_\ell, \dots, r_{n+1})$  en accepterende følge for  $M$ , og ordet den genkender er  $xy^iz$ .

**Eksempel 1.74:** Sproget

$C = \{w \mid \text{antallet af 0 i } w \text{ er lig med antallet af 1}\} \subseteq \{0, 1\}^*$  er ikke regulært.

(Samme bevis som for eksempel 1.73)

**Bemærkning** (opgave 1.48): Sproget

$D = \{w \mid \text{antallet af 01 i } w \text{ er lig med antallet af 10}\} \subseteq \{0, 1\}^*$  er regulært!

(Men *kun* over alfabetet  $\{0, 1\}$ ; hvis alfabetet f.x. er  $\{0, 1, 2\}$ , er  $D$  ikke regulært ... )

**Bevis:**

