

Regulære udtryk

Definition af regulære udtryk (en formalisme der beskriver regulære sprog)

Syntax		Semantik
$r = \emptyset$	\rightarrow	$L(\emptyset) = \emptyset$
$r = a \in \Sigma$	\rightarrow	$L(a) = \{a\}$
$r = r_1 + r_2$	\rightarrow	$L(r_1 + r_2) = L(r_1) \cup L(r_2)$
$r = r_1 r_2$	\rightarrow	$L(r_1 r_2) = L(r_1) L(r_2)$
$r = r_1^*$	\rightarrow	$L(r_1^*) = L(r_1)^*$

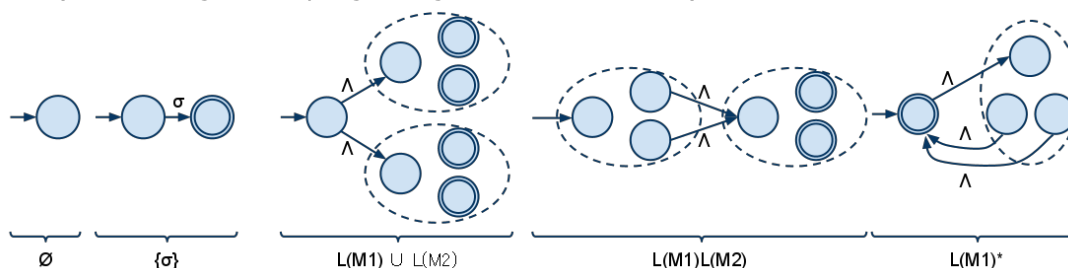
Forbindelse til FA'er (deklarative vs operationelle)

FA $M = \{Q, \Sigma, q_0, A, \delta\}$

Definer $\delta: Q \times \Sigma \rightarrow Q$

Har samme udtrykskraft som regulære udtryk (Bevis)

Kleene 1 (Ethvert regulært sprog kan genkendes af en FA)



Kleene 2 (For enhver FA M er sproget $L(M)$ regulært)

Introducer $L(p, q) = \{x \mid \delta^*(p, x) = q\}$

Hvis vi kan vise at $L(p, q)$ er regulært så er $L(M)$ regulært da,

$L(M) = \bigcup \{L(q_0, q) \mid q \in A\}$

Introducer $L(p, q, k)$

x går igennem en tilstand r hvis $x = yz$ hvor $\delta^*(p, y) = r$ og $\delta^*(r, z) = q$

Numerer states fra 1 til n

Vis at $L(M)$ er regulært

Base case $k = 0$

$L(p, q, 0) = \{\sigma \in \Sigma \mid \delta(p, \sigma) = q\}$

I.H. $L(p, q, k)$ er regulært for alle $p, q \in Q$

Induktion vis at $L(p, q, k+1)$ er regulært (**tegn**)

$L(p, q, k+1) = L(p, q, k) \cup L(p, k+1, k) L(k+1, k+1, k)^* L(k+1, q, k)$

De enkelte sprog er regulære jvf. induktions hypotesen, og resultatet er regulært da de regulære sprog er lukket under

forening, konkatenering og kleene stjerne

Endelige automater

Definition af FA'er

Regulære udtryk bruges til at specificere regulære sprog, mens endelige automater bruges til at afgøre om en given streng er indeholdt i et sprog.

Regulære udtryk og endelige automater har samme udtrykskraft

Beskrives med en 5-tupel

Produktkonstruktion (sproget for en FA er lukket under forening, snit og differens)

$$L(M) = L(M_1) \cup L(M_2)$$

$$M = \{Q, \Sigma, q_0, A, \delta\}$$

$$Q = Q_1 \times Q_2 \text{ (en state repræsenterer et par af states)}$$

$$q_0 = (q_1, q_2)$$

$$A = \{(p, q) \mid p \in A_1 \vee q \in A_2\}$$

$$\delta((p, q), \sigma) = (\delta_1(p, \sigma), \delta_2(q, \sigma))$$

$$\delta^*((p, q), x) = (\delta_1^*(p, x), \delta_2^*(q, x)) \text{ (bevis ved induktion i længden af } x \text{)}$$

$$\text{Basis } x = \Lambda \Rightarrow \delta^*((p, q), \Lambda) = (p, q)$$

$$\delta^*((p, q), \Lambda) = (\delta_1^*(p, \Lambda), \delta_2^*(q, \Lambda))$$

$$\text{I.H. Det gælder for } |x| < n$$

$$\text{Vis Det gælder for } n = n+1$$

$$\text{Bevis } x = y\sigma$$

$$\begin{aligned} \delta^*((p, q), y\sigma) &\Rightarrow \delta(\delta^*((p, q), y), \sigma) \Rightarrow \delta(\delta_1^*(p, y), \delta_2^*(q, y)), \sigma) \Rightarrow \\ &(\delta_1(\delta_1^*(p, y)), \delta_2(\delta_2^*(q, y)), \sigma) \Rightarrow (\delta_1^*(p, y\sigma), \delta_2^*(q, y\sigma)) \end{aligned}$$

Bevis

$$\begin{aligned} x \in L(M) &\Leftrightarrow \delta^*(q_0, x) \in A \Rightarrow \delta^*((q_1, q_2), x) \in A \Rightarrow (\delta_1^*(p, x), \delta_2^*(q, x)) \in A \Rightarrow \\ &(\delta_1^*(p, x) \in A_1 \vee \delta_2^*(q, x) \in A_2) \Rightarrow x \in L(M_1) \vee x \in L(M_2) \Rightarrow x \in L(M_1) \cup L(M_2) \end{aligned}$$

$$L(M_1) \cup L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \vee q \in A_2\}$$

$$L(M_1) \cap L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \wedge q \in A_2\}$$

$$L(M_1) - L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \wedge q \notin A_2\}$$

Vis L ikke er regulært vha. **kontraponering**

Antag $A, C \in \Sigma^*$ hvor A er regulært og C ikke er, og $A \cup B = C$, så er B ikke regulær.

Lukketsegenskaber

Produktkonstruktion

$$L(M) = L(M_1) \cup L(M_2)$$

$$M = \{Q, \Sigma, q_0, A, \delta\}$$

$$Q = Q_1 \times Q_2 \text{ (en state repræsenterer et par af states)}$$

$$q_0 = (q_1, q_2)$$

$$A = \{(p, q) \mid p \in A_1 \vee q \in A_2\}$$

$$\delta((p, q), \sigma) = (\delta_1(p, \sigma), \delta_2(q, \sigma))$$

$$\delta^*((p, q), x) = (\delta_1^*(p, x), \delta_2^*(q, x)) \text{ (bevis ved induktion i længden af } x)$$

$$\text{Basis } x = \Lambda \Rightarrow \delta^*((p, q), \Lambda) = (p, q)$$

$$\delta^*((p, q), \Lambda) = (\delta_1^*(p, \Lambda), \delta_2^*(q, \Lambda))$$

$$\text{I.H. Det gælder for } |x| < n$$

$$\text{Vis Det gælder for } n = n+1$$

$$\text{Bevis } x = y\sigma$$

$$\begin{aligned} \delta^*((p, q), y\sigma) &\Rightarrow \delta(\delta^*((p, q), y), \sigma) \Rightarrow \delta(\delta_1^*(p, y), \delta_2^*(q, y)), \sigma) \Rightarrow \\ &(\delta_1(\delta_1^*(p, y)), \delta_2(\delta_2^*(q, y)), \sigma) \Rightarrow (\delta_1^*(p, y\sigma), \delta_2^*(q, y\sigma)) \end{aligned}$$

Bevis

$$\begin{aligned} x \in L(M) &\Leftrightarrow \delta^*(q_0, x) \in A \Rightarrow \delta^*((q_1, q_2), x) \in A \Rightarrow (\delta_1^*(p, x), \delta_2^*(q, x)) \in A \Rightarrow \\ &(\delta_1^*(p, x) \in A_1 \vee \delta_2^*(q, x) \in A_2) \Rightarrow x \in L(M_1) \vee x \in L(M_2) \Rightarrow x \in L(M_1) \cup L(M_2) \end{aligned}$$

$$L(M_1) \cup L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \vee q \in A_2\}$$

$$L(M_1) \cap L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \wedge q \in A_2\}$$

$$L(M_1) - L(M_2) \Rightarrow A = \{(p, q) \mid p \in A_1 \wedge q \notin A_2\}$$

Vis L ikke er regulært vha. **kontraponering**

Antag $A, C \in \Sigma^*$ hvor A er regulært og C ikke er, og $A \cup B = C$, så er B ikke regulær.

Kontekstfri sprog - ikke lukket under fællesmængde og komplement

$$L = \{a^i b^j c^k \mid i < j, i < k\} \notin CFL$$

$$L_1 = \{a^i b^j c^k \mid i < j\}$$

$$L_2 = \{a^i b^j c^k \mid i < k\}$$

$$L_1 \cap L_2 = L \Rightarrow CFG \text{ er ikke lukket under fællesmængde}$$

Nondeterministiske automater

Definition af NFA (5-tupel, NFA vs FA, beskriver regulære sprog)

$$M = \{Q, \Sigma, q_0, A, \delta\}$$

$\delta: Q \times \Sigma \rightarrow 2^Q$ et symbol kan gå til mange states

Vigtigt

$$\delta^*(q, x) = \{q\} \text{ hvis } x = \Lambda$$

(uden lambdatransitioner)

$$\delta^*(q, y\sigma) = \bigcup \{\delta(r, \sigma) \mid r \in \delta^*(q, y)\}$$

(uden lambdatransitioner)

Sproget for en NFA

$$x \in L(M) \text{ hvis } \delta^*(q_0, x) \cap A \neq \emptyset$$

Determinisering

$$Q_1 = 2^Q \text{ (en tilstand repræsenterer en mængde af tilstande)}$$

$$q_1 = \{q_0\}$$

$$\delta_1(q, \sigma) = \bigcup \{\delta(p, \sigma) \mid p \in q\}$$

$$A_1 = \{q \in Q_1 \mid q \cap A \neq \emptyset\}$$

$$\text{Bevis } \delta_1^*(q_1, x) = \delta^*(q_0, x)$$

$$\text{Base case } x = \Lambda$$

$$\delta_1^*(q_1, \Lambda) = q_1 = \{q_0\} = \delta^*(q_0, \Lambda)$$

$$\text{I.H. } \delta_1^*(q_1, x) = \delta^*(q_0, x)$$

$$\text{Vis } x = y\sigma$$

$$\delta_1^*(q_1, y\sigma) =$$

$$\delta_1(\delta_1^*(q_1, y), \sigma) =$$

$$\delta_1(\delta^*(q_0, y), \sigma) =$$

$$\bigcup \{\delta(p, \sigma) \mid p \in \delta^*(q_0, y)\} =$$

$$\delta^*(q_0, y\sigma)$$

Minimering af automater

Definition af FA'er og deres sprog

En FA kan afgøre om en streng tilhører et givent regulært sprog

$$M = \{Q, \Sigma, q_0, A, \delta\}$$

Definer udvidede transitionsfunktion

$$\delta^*(q, xa) = \delta(\delta^*(q, x), a)$$

Minimeringsalgoritmen + bevis for korrekthed

Minimal FA $M = \{Q, \Sigma, q_0, A, \delta\}$

$Q = Q_L$ hvor Q_L er ækvivalensklasserne for I_L

$$q_0 = [\lambda]$$

$$A = \{q \mid q \subseteq L\}$$

$$\delta([x], a) = [xa], \text{ dette gælder da } [x] I_L [y] \Rightarrow [xa] I_L [ya]$$

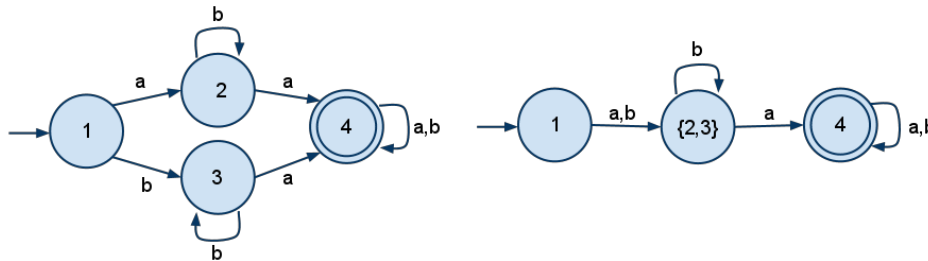
$$p \neq q \Leftrightarrow (p, q) \in S$$

1. Fjern alle uopnåelige tilstande

$$2. (p \in A \wedge q \notin A) \vee (p \notin A \wedge q \in A) \Rightarrow (p, q) \in S$$

$$3. (\exists \sigma \in \Sigma: (\delta(p, \sigma) = r, \delta(q, \sigma) = s) \in S) \Rightarrow (p, q) \in S$$

Hvis z er vidne på at enhver streng i L_r er skelnelig fra enhver streng i L_s , så er σz et vidne på at enhver streng i L_p er skelnelig fra enhver streng i L_q



Begrænsninger af regulære sprog

Tag udgangspunkt i A^nB^n (er sproget regulært?)

Intuitivt bevis (fast bevis kan laves vha. Pumping-lemmaet)

"Pumping"-lemmaet for regulære sprog

Antag der findes en streng x , hvor $|x| \geq |Q|$ så besøges samme tilstand flere gange
Tegn tegning med u , v og w

1. $|v| > 0$
2. $|uv| \leq n$
3. $\forall m \geq 0, uv^mw \in L$

Pumping-lemmaet

$\exists n > 0$

$\forall x \in L$, hvor $|x| > n$

$\exists u, v, w \in \Sigma^*$, hvor $x = uvw$

$\forall m \geq 0: uv^mw \in L$

Bevis at et sprog ikke er regulært (kontraponer Pumping-lemmaet og nå modstrid)

$\forall n > 0$

$\exists x \in L$, hvor $|x| > n$

$\forall u, v, w \in \Sigma^*$, hvor $x = uvw$

$\exists m \geq 0: uv^mw \notin L$, så er L ikke regulært

Eksempel på brug af Pumping-lemmaet (A^nB^n)

Kontekstfri grammatikker

Definition af CFG'er

$G = \{V, \Sigma, S, P\}$ variable, terminaler, startvariabel, produktioner

$S \in V$

$P: A \rightarrow \alpha$, hvor $A \in V$ og $\alpha \in \{V \cup \Sigma\}^*$

Opskriv $S \rightarrow aSb \mid \Lambda$, forklar hvad en derivation er

$L(G) = \{x \in \Sigma^* \mid S \Rightarrow^* x\}$

Et sprog L er kontekstfrit, hvis der findes en grammatik så $L = L(G)$

Find ud af om et sprog ikke er kontekstfrit

Chomsky Normal Form

Enhver produktion er på følgende form

$A \rightarrow BC$, hvor $B, C \in V$

$A \rightarrow \sigma$, hvor $\sigma \in \Sigma$

Theorem siger: for enhver CGF G findes en CFG G_1 på CNF så $L(G) = L(G_1) - \{\Lambda\}$

"Pumping"-lemmaet for CFG'er

Antag L er et kontekstfrit sprog. Så findes der et tal n , så for alle $u \in L$, hvor $|u| \geq n$, kan u skrives som $u = vwxyz$

1. $|wy| > 0$
2. $|wxy| \leq n$
3. for alle $m \geq 0$, $vw^mxy^mz \in L$

For en tilstrækkelig lang derivation vil en variabel derivere til sig selv så

$S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vwxyz$

$S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vw^2Ay^2z \Rightarrow^* vw^3Ay^3z \Rightarrow^* \dots$

Oversæt til Chomsky Normal Form

Binært derivationstræ

Binært træ indeholder max 2^h blade (terminalsymboler)

Sæt p til antallet af variable i grammatikken

Lad n være 2^{p+1}

Hvis $|u| \geq n$ må træet være mindst $p+1$ højt

Derfor må den samme variabel gå igen mindst 2 gange, deriverer til sig selv

Derived opnåes tilstanden $vAz \Rightarrow^* vwAyz \Rightarrow^* vw^2Ay^2z \Rightarrow^* vw^3Ay^3z \Rightarrow^* \dots$

Vis på derivationstræet hvor $vwxyz$ kommer fra

Vis hvorfor $|wy| > 0$ er sandt (pga. Chomsky)

Vis hvorfor $|wxy| \leq n$ er sandt (pga. $n = 2^{p+1}$)

Vis hvorfor $vw^mxy^mz \in L$ (indsæt flere af det øverste A (eller fjern det))