

Syntaks og semantik

Lektion 7

4 marts 2008

Forord

- 1 Pushdown-automater
- 2 Automater med stæcke
- 3 Grammatikker
- 4 Chomsky-hierarkiet

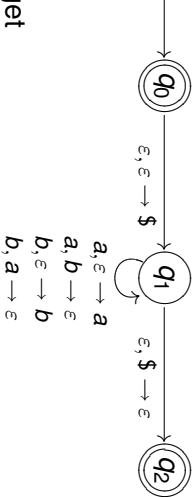
Definition 2.13: En **pushdown-automat (PDA)** er en 6-tupel $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$, hvor delene er

- 1 Q : en endelig mængde af tilstande
- 2 Σ : input-alfabetet
- 3 Γ : stack-alfabetet
- 4 $\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon \rightarrow \mathcal{P}(Q \times \Sigma_\epsilon \times \Gamma_\epsilon)$: transitionsfunktionen
- 5 $q_0 \in Q$: starttilstanden
- 6 $F \subseteq Q$: mængden af accepttilstande

M siges at **acceptiere** et ord $w \in \Sigma^*$ hvis der findes $m \in \mathbb{N}$ og $w_1, w_2, \dots, w_m \in \Sigma_\epsilon, r_0, r_1, \dots, r_m \in Q$ og $s_0, s_1, \dots, s_m \in \Gamma^*$ således at $w = w_1 w_2 \dots w_m$ og

- 1 $r_0 = q_0$ og $s_0 = \epsilon$,
- 2 for alle $i = 0, 1, \dots, m - 1$ findes $a, b \in \Gamma_\epsilon$ og $t \in \Gamma^*$ som opfylder $s_i = at, s_{i+1} = bt$ og (r_i, w_{i+1}, a) , og
- 3 $r_m \in F$.

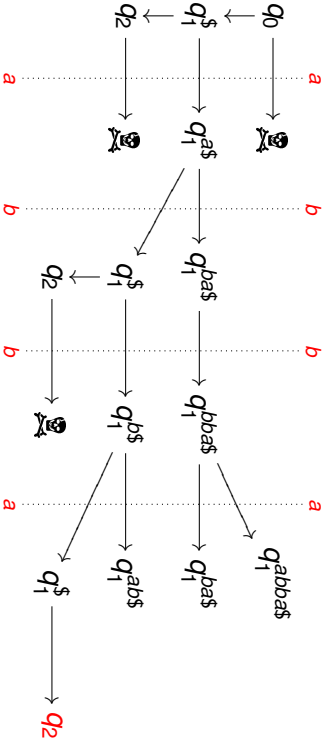
Eksempel:



Genkender sproget

$\{w \in \{a, b\}^* \mid \text{antallet af } a \text{ i } w = \text{antallet af } b \text{ i } w\}$

At læse strengen abba:



Definition: Et sprog siges at være **kontekstfrit** hvis der findes en CFG der genererer det.

Sætning 2.20: Et sprog er kontekstfrit hvis og kun hvis der findes en PDA der genkender det.

- at bevise **CFG \Rightarrow PDA**:
Lav en CFG G om til en ("generaliseret") PDA med 3 tilstande:
Fra q_{start} til q_{loop} pushes startsymbolet fra G på stæcken. Fra q_{loop} til sigeslv er der transitioner der
 - ekspanderer en variabel i G til en af dens højresider i G s produktioner,
 - forsøger at matche en terminal fra input med en terminal fra stæcken.
- Fra q_{loop} til q_{accept} er der en transition der kun er tændt når stæcken er tom.
- at bevise **PDA \Rightarrow CFG**: Senere i dag

5 / 17

Definition: En **automat med k stæcke**, for $k \in \mathbb{N}_0$, er en 6-tupel $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$, hvor delene er

- 1. Q : en endelig mængde af tilstande
- 2. Σ : input-alfabetet
- 3. Γ : stack-alfabetet
- 4. $\delta : Q \times \Sigma_{\epsilon} \times \Gamma_{\epsilon}^k \rightarrow \mathcal{P}(Q \times \Gamma_{\epsilon}^k)$: transitionfunktionen
- 5. $q_0 \in Q$: starttilstanden
- 6. $F \subseteq Q$: mængden af accepttilstande

- $k = 0$: **NFA**
- $k = 1$: **PDA**
- **$k \geq 2$: Turing-maskine!**
 - to stæcke er nok!

6 / 17

Definition: En **grammatik** er en 4-tupel $G = (V, \Sigma, R, S)$, hvor delene er

- 1. V : en endelig mængde af variable
- 2. Σ : en endelig mængde af terminaler, med $V \cap \Sigma = \emptyset$
- 3. $R : (V \cup \Sigma)^+ \rightarrow \mathcal{P}((V \cup \Sigma)^*)$: produktioner
- 4. $S \in V$: startvariablen
- alle produktioner på formen $A \rightarrow \epsilon$, $A \rightarrow a$ eller $A \rightarrow aB$, for $A, B \in V$ og $a \in \Sigma$: **regulær** grammatik
- alle produktioner på formen $A \rightarrow w$, for $A \in V$ og $w \in (V \cup \Sigma)^*$: **kontekstfri** grammatik
- alle produktioner på formen $uAv \rightarrow uvw$, for $A \in V$ og $u, v, w \in (V \cup \Sigma)^*$: **kontekst-sensitiv** grammatik

Eksempel på en kontekst-sensitiv grammatik:

$$S \rightarrow aBSc \mid abc \quad Ba \rightarrow aB \quad Bb \rightarrow bb$$

Genererer sproget $\{a^n b^n c^n \mid n \in \mathbb{N}_+\}$

7 / 17

	Type 3	Type 2	Type 1	Type 0
	regulære sprog	kontekstfrie sprog	kontekst-sensitive sprog	rekursivt enumerable sprog
	regulære grammatikker	kontekstfrie grammatikker	kontekst-sensitive grammatikker	generelle grammatikker
endelige automater		pushdown-automater	lineært begrænsede Turing-maskiner	Turing-maskiner
determinisme	ingen indskrænkning	indskrænkning	vides ikke	ingen indskrænkning
lukning: $\cup, \circ, *$	ja	ja	ja	ja
\cap	ja	nej	ja	ja
-	ja	nej	ja	nej

8 / 17

Kontekstfri og ikke kontekstfri sprog

- 5 Ethvert sprog genkendt af en PDA er kontekstfrit
- 6 Pumpeløkke for kontekstfri sprog

9 / 17

PDA \Rightarrow CFG

Pumpeløkke

Lemma 2.27: Lad Σ være et alfabet og P en PDA over Σ . Da findes en CFG G over Σ med $\llbracket G \rrbracket = \llbracket P \rrbracket$.

Bevis: Lad $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$. Vi konstruerer $G = (V, \Sigma, R, S)$:

- 1 Sørg for at P kun har én accepttilstand q_a og at stacken tømmes før P går i q_a .
- 2 Lad $V = \{A_{pq} \mid p, q \in Q\}$, og sørg for at A_{pq} deriverer præcist de strenge som bringer P fra p med tom stack til q med tom stack.
- 3 Lad $S = A_{q_0 q_a}$. Voilà!

PDA \Rightarrow CFG

Pumpeløkke

Lemma 2.27: Lad Σ være et alfabet og P en PDA over Σ . Da findes en CFG G over Σ med $\llbracket G \rrbracket = \llbracket P \rrbracket$.

Bevis: Lad $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$. Vi konstruerer $G = (V, \Sigma, R, S)$:

- 1 Sørg for at P kun har én accepttilstand q_a og at stacken tømmes før P går i q_a .

Nyt stacksymbol $\$$. Tre nye tilstande: q_s, q_e og q_a . Nye

transitioner: $q_s \xrightarrow{\epsilon, \epsilon \rightarrow \$} q_0, q \xrightarrow{\epsilon, \epsilon \rightarrow \epsilon} q_e$ for alle $q \in F$,

$q_e \xrightarrow{\epsilon, a \rightarrow \epsilon} q_e$ for alle $a \in \Sigma$, og $q_e \xrightarrow{\epsilon, \$ \rightarrow \epsilon} q_a$.

Sørg for at enhver transition enten pusher eller popper.

- Erstat enhver transition $q \xrightarrow{a, b \rightarrow c} r$ med $q \xrightarrow{a, b \rightarrow \epsilon} q_1 \xrightarrow{\epsilon, \epsilon \rightarrow c} r$
- Erstat enhver transition $q \xrightarrow{a, \epsilon \rightarrow \epsilon} r$ med $q \xrightarrow{a, \epsilon \rightarrow X} q_1 \xrightarrow{\epsilon, X \rightarrow \epsilon} r$ for et eller andet symbol $x \in \Gamma$.

11 / 17

PDA \Rightarrow CFG

Pumpeløkke

Lemma 2.27: Lad Σ være et alfabet og P en PDA over Σ . Da findes en CFG G over Σ med $\llbracket G \rrbracket = \llbracket P \rrbracket$.

Bevis: Lad $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$. Vi konstruerer $G = (V, \Sigma, R, S)$:

- 2 Lad $V = \{A_{pq} \mid p, q \in Q\}$, og sørg for at A_{pq} deriverer præcist de strenge som bringer P fra p med tom stack til q med tom stack.
 - Lav en produktion $A_{pp} \rightarrow \epsilon$ for alle $p \in Q$ (terminering)
 - Lav en produktion $A_{pq} \rightarrow A_{pr}A_{rq}$ for alle $p, q, r \in Q$ (rekursion)
 - For alle $p, q, r, s \in Q$: Hvis $p \xrightarrow{a, \epsilon \rightarrow t} r$ og $s \xrightarrow{b, t \rightarrow \epsilon} q$ for nogle $a, b \in \Sigma_\epsilon$ og et $t \in \Gamma$: Lav en produktion $A_{pq} \rightarrow aA_{rs}b$ (produktion)

– der skal argumenteres for at dette giver det rigtige resultat!

Sætning 2.34: For ethvert kontekstfrit sprog A findes der et (naturligt) tal p således at ethvert ord $s \in A$ der har længde mindst p kan opsplittes i fem stykker, $s = uvxyz$, med

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

Anvendelse: Vis a sproget X ikke er kontekstfrit:

Antag at X er kontekstfrit. Så må det opfylde pumpelemmaet. Lad p være pumpe længden.

Find en streng s som

- har $|s| \geq p$, dvs. *bør kunne pumpes*,
- men som *ikke kan pumpes*, ligegyldigt hvordan man opsplitter $s = uvxyz$.

Modstrid!

13 / 17

Bevis: Lad $G = (V, \Sigma, P, S)$ være en CFG med $\|G\| = A$.

- 1 Lad b være længden af den længste streng på højresiden af produktionerne i G : $b = \max\{|s| \mid s \in P(A), A \in V\}$
- 2 Lad $p = b^{|V|+1}$. **Fejl i bogen!** Tag et $s \in A$ med $|s| \geq p$. $|V|$ er antallet af variable i G .

Bevis: Lad $G = (V, \Sigma, P, S)$ være en CFG med $\|G\| = A$.

- 1 Lad b være længden af den længste streng på højresiden af produktionerne i G : $b = \max\{|s| \mid s \in P(A), A \in V\}$
- 2 Lad $p = b^{|V|+1}$. **Fejl i bogen!** Tag et $s \in A$ med $|s| \geq p$.
- 3 Lad τ være et af de parsetræer for s der har færrest punkter. τ har **højde mindst $|V| + 1$** .
Lad h være højden af τ . Hvert punkt i τ har **højst b sønner**, så τ har **højst b^h blade**. Tegnene i s står i bladene, så s har længde højst b^h . Men $|s| > b^{|V|}$, så $h > |V|$.

15 / 17

Bevis: Lad $G = (V, \Sigma, P, S)$ være en CFG med $\|G\| = A$.

- 1 Lad b være længden af den længste streng på højresiden af produktionerne i G : $b = \max\{|s| \mid s \in P(A), A \in V\}$
- 2 Lad $p = b^{|V|+1}$. **Fejl i bogen!** Tag et $s \in A$ med $|s| \geq p$.
- 3 Lad τ være et af de parsetræer for s der har færrest punkter. τ har **højde mindst $|V| + 1$** .
- 4 Lad ℓ være en sti i τ af længde mindst $|V| + 2$.
- 5 ℓ indeholder mindst $|V| + 1$ variable (og én terminal), så blandt de sidste $|V| + 1$ variable i ℓ er der en der forekommer to gange. Kald den R .
- 6 Lad x være den delstreng af s der deriveres af den sidste forekomst af R . Strengen der deriveres af den næstsidste forekomst af R kan da skrives vxy , og $s = uvxyz$.
Dvs. $R \Rightarrow^* x$, $R \Rightarrow^* vRy \Rightarrow^* vxy$, og $S \Rightarrow^* uRz \Rightarrow^* uvRyz \Rightarrow^* uvxyz$.

14 / 17

16 / 17

- 6 Lad x være den delstreng af s der derivatives af den *sidste* forekomst af R . Strengen der derivatives af den *næstsidste* forekomst af R kan da skrives vxy , og $s = uvxyz$.
- 7 Den næstsidste forekomst af R er blandt de sidste $|V| + 1$ variable i ℓ , så deltræet med dette R som rod har højde *højst* $|V| + 1$, så $|vxy| \leq b^{|V|+1} = p$. **Fejl i bogen!**
- 8 Ved at erstatte deltræet med det *næstsidste* R som rod, med deltræet med det *sidste* R som rod fås derivationen $S \xRightarrow{*} uRz \xRightarrow{*} uxz$. Dvs.
 - $uxz = uv^0xy^0z \in A$
 - $|vy| > 0$, for ellers ville $s = uxz$, og det parsetræ for uxz vi lige har lavet er mindre end det vi startede med. Modstrid til (3).
- 9 Ved at erstatte deltræet med det *sidste* R som rod, med deltræet med det *næstsidste* R som rod fås derivationen $S \xRightarrow{*} uRz \xRightarrow{*} uvRyz \xRightarrow{*} uv^2Ry^2z \xRightarrow{*} uv^2xy^2z$.
Ved at gentage dette fås derivationen til $uv^i xy^i z$ for alle $i \in \mathbb{N}$.