

Syntaks og semantik

Lektion 8

13 marts 2007

Perspektivering

- 1 Automater med stacke
- 2 Grammatikker
- 3 Chomsky-hierarkiet

Definition: En **automat med k stacke**, 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 $\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon^k \rightarrow \mathcal{P}(Q \times \Gamma_\epsilon^k)$: transitionsfunktionen
 - 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 stacke er nok!

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 $R : (V \cup \Sigma)^* V (V \cup \Sigma)^* \rightarrow \mathcal{P}((V \cup \Sigma)^*)$: produktioner
 - 4 $S \in V$: startvariablen
- alle produktioner på formen $A \rightarrow w$, for $A \in V$ og $w \in (V \cup \Sigma)^*$: **kontekstfri** grammatik
 - alle produktioner på formen $A \rightarrow \varepsilon$, $A \rightarrow a$ eller $A \rightarrow aB$, for $A, B \in V$ og $a \in \Sigma$: **regulær** grammatik

Eksempel på en ikke-kontekstfri grammatik:

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

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

	Type 3	Type 2	Type 0
	regulære sprog	kontekstfrie sprog	<i>rekursivt enumerable</i> sprog
	regulære grammatikker	kontekstfrie grammatikker	generelle grammatikker
	endelige automater	pushdown-automater	Turing-maskiner
determinisme lukket under:	ingen indskrænkning	indskrænkning	ingen indskrænkning
$\cup, \circ, *$	ja	ja	ja
\cap	ja	nej	ja
-	ja	nej	nej

Ikke-kontekstfrie sprog

- 4 Pumpelemmaet for kontekstfrie sprog
- 5 Et par indirekte beviser
- 6 Ikke-kontekstfrie sprog

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 at sproget X ikke er kontekstfrit:

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

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!

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

- 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 $\llbracket G \rrbracket = A$.

- 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 $h > |V|$.

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

- 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 leveres af den *sidste* forekomst af R . Strengen der leveres af den *næstsidste* forekomst af R kan da skrives vxy , og $s = uvxyz$.

Dvs. $R \xRightarrow{*} x$, $R \xRightarrow{*} vRy \xRightarrow{*} vxy$, og

$S \xRightarrow{*} uRz \xRightarrow{*} uvRyz \xRightarrow{*} uvxyz$.

- 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 $s = uvxyz$.
- 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 derivationer til $uv^i xy^i z$ for alle $i \in \mathbb{N}$.

Sætning: $\sqrt{2}$ er et irrationelt tal.

Bevis:

- ① *Antag* at $\sqrt{2}$ er et *rationelt* tal.
- ② Så må det kunne skrives som en brøk: $\sqrt{2} = \frac{a}{b}$, for to positive heltal a og b .
- ③ Lad brøken være *reduceret*, dvs. specielt er ikke både a og b lige tal.
- ④ $\frac{a}{b} = \sqrt{2}$ medfører at $2b^2 = a^2$.
- ⑤ Hvis a er ulige, er a^2 også ulige, **modstrid** til (4).
- ⑥ Dvs. a må være et lige tal, og med (3) må b så være ulige.
- ⑦ Skriv $a = 2c$. Så er $2b^2 = a^2 = 4c^2$, dvs. $b^2 = 2c^2$.
- ⑧ Men b er ulige, så det er b^2 også, **modstrid** til (7).
- ⑨ Antagelsen om at $\sqrt{2}$ var et rationelt tal ledte frem til et modstrid, så den må være forkert. Konklusion: $\sqrt{2}$ er et irrationelt tal.

Sætning: Der findes uendeligt mange primtal.

Bevis:

- ① *Antag* at der kun findes endeligt mange primtal. Kald dem p_1, p_2, \dots, p_k .
- ② Lad $N = p_1 p_2 \dots p_k + 1$.
- ③ N er større end ethvert af primtallene, så det kan ikke være et primtal selv.
- ④ Dvs. der er et primtal der går op i N . Kald det p_i .
- ⑤ Men $N - 1 = p_1 p_2 \dots p_k$, så p_i går også op i $N - 1$.
- ⑥ Derfor går p_i op i $N - (N - 1) = 1$, **modstrid**.
- ⑦ Antagelsen om at der kun findes endeligt mange primtal ledte frem til et modstrid, så den må være forkert.
Konklusion: Der findes uendeligt mange primtal. Euklid havde ret!

Eksempel 2.36: Sproget $B = \{a^n b^n c^n \mid n \in \mathbb{N}_0\}$ er ikke kontekstfrit:

Bevis:

- ❶ Antag at B er kontekstfrit, og lad p være dets pumpelængde.
- ❷ Lad $s = a^p b^p c^p$. (*Et smart valg!*) Vi har $|s| \geq p$.
- ❸ Lad $s = uvxyz$ være den opsplitning af s som pumpelemmaet garanterer. (*Vi ved den findes. Vi ved ikke hvordan den ser ud!*)
- ❹ Hvis v og y hver kun indeholder én slags af symbolerne a , b og c , er der et af symbolerne der ikke er med i v eller y . Strengen uv^2xy^2z indeholder så for få symboler af denne slags og er derfor ikke indeholdt i B , **modstrid!**
- ❺ Hvis v eller y indeholder mere end én slags symboler, optræder de i uv^2xy^2z i forkert rækkefølge
 $\Rightarrow uv^2xy^2z \notin B$, **modstrid!**
- ❻ Ligegyldigt hvad får vi en modstrid. \Rightarrow antagelsen forkert
 $\Rightarrow B$ er ikke kontekstfrit.

Eksempel 2.38: Sproget $D = \{ww \mid w \in \{0, 1\}^*\}$ er ikke kontekstfrit:

Bevis:

- 1 Antag at D er kontekstfrit, og lad p være dets pumpelængde.
- 2 Lad $s = 0^p 1^p 0^p 1^p$. Vi har $|s| \geq p$. Lad $s = uvxyz$ være den opsplitning af s som pumpelemmaet garanterer.
- 3 Hvis strengen vxy er en del af det første $0^p 1^p$ i s , starter anden halvdel af uv^2xy^2z med et 1. Men første halvdel starter stadig med 0, så $uv^2xy^2z \notin D$, **modstrid!**
- 4 Hvis strengen vxy er en del af det andet $0^p 1^p$ i s , slutter første halvdel af uv^2xy^2z med et 0, men anden halvdel slutter med 1, så $uv^2xy^2z \notin D$, **modstrid!**
- 5 Så strengen vxy må indeholde midten af s , dvs. vxy er en del af det midterste $1^p 0^p$. Men $|vy| > 0$, så $|x| < |vxy|$, dvs. $uv^0xy^0z = 0^p 1^i 0^j 1^p$ med $i < p$ eller $j < p$, så $uv^0xy^0z \notin D$, **modstrid!**