Syntaks og semantik

Lektion 8

13 marts 2007

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. $\delta : Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon^k \to \mathcal{P}(Q \times \Gamma_\varepsilon^k)$: 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!

Perspektivering

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

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. $R : (V \cup \Sigma)^* V (V \cup \Sigma)^* \to \mathcal{P}\big((V \cup \Sigma)^*\big)$: produktioner
4. $S \in V$: startvariablen

- alle produktioner på formen $A \to w$, for $A \in V$ og $w \in (V \cup \Sigma)^*$: kontekstfri grammatik
- alle produktioner på formen $A \to \varepsilon$, $A \to a$ eller $A \to aB$, for $A, B \in V$ og $a \in \Sigma$: regulær grammatik

Eksempel på en ikke-kontekstfri grammatik:

$$S \to aBSc \mid abc \qquad Ba \to aB \qquad Bb \to bb$$

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

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

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 $|vxy| \le p$,
- 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 pumpelængden.

Find en streng s som

- har $|s| \ge p$, 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 $[\![G]\!] = A$.

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 $|s| \ge p$. $|V|$ 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 $h > |V|$.

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 $S \overset{*}{\Rightarrow} uRz \overset{*}{\Rightarrow} uxz$. 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 $S \overset{*}{\Rightarrow} uRz \overset{*}{\Rightarrow} uvRyz \overset{*}{\Rightarrow} uv^2Ry^2z \overset{*}{\Rightarrow} uv^2xy^2z$.
   Ved at gentage dette fås derivationer til uv^ixy^iz for alle $i \in \mathbb{N}$.

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. $R \overset{*}{\Rightarrow} x$, $R \overset{*}{\Rightarrow} vRy \overset{*}{\Rightarrow} vxy$, og $S \overset{*}{\Rightarrow} uRz \overset{*}{\Rightarrow} uvRyz \overset{*}{\Rightarrow} uvxyz$.

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 $p_1, p_2, \ldots, p_k$.
2. Lad $N = p_1 p_2 \ldots p_k + 1$.
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 $N - 1 = p_1 p_2 \ldots p_k$, 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:

1. *Antag* at B er kontekstfrit, og lad p være dets pumpelængde.
2. Lad $s = a^p b^p c^p$. *(Et smart valg!)* Vi har $|s| \geq p$.
3. Lad $s = uvxyz$ være den opsplitning af s som pumpelemmaet garanterer. *(Vi ved den findes. Vi ved ikke hvordan den ser ud!)*
4. 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!
5. 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!
6. 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!