TENTAMEN

Basismodellen in de Informatica

vakcode: 211180
datum: 26 juni 2008
tijd: 9:00–12:30 uur

VOORBEELDUITWERKING

Algemeen

- Bij dit tentamen mag gebruik worden gemaakt van het boek van Sudkamp, van de handleiding van Basismodellen in de Informatica, en van de handouts van de colleges.
- Dit tentamen bestaat uit 6 opgaven, waarvoor in het totaal 90 punten behaald kunnen worden. Het minimale aantal punten per opgave bedraagt 0. Het cijfer is het aantal punten gedeeld door 10, afgerond tot een geheel getal, plus 1.

Opgave 1 **(20 punten)**

Gegeven is de volgende NFA-λ:

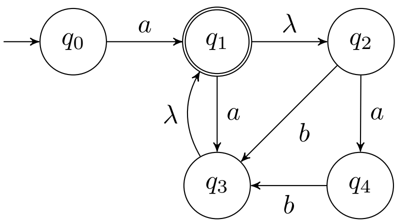
$$
\begin{aligned}
M &= \big(\,\{q_0, q_1, q_2, q_3, q_4\}\ ,\ \{a, b\}\ ,\ \delta\ ,\ q_0\ ,\ \{q_1\}\,\big),\ \text{waarbij} \\
\delta &= \{(q_0, a, q_1), (q_1, \lambda, q_2), (q_2, a, q_4), (q_2, b, q_3), (q_3, \lambda, q_1), (q_1, a, q_3), (q_4, b, q_3)\}
\end{aligned}
$$

a. Teken het toestandsdiagram van M.

b. Beschrijf $\mathcal{L}(M)$, de taal die door M geaccepteerd wordt, door middel van een reguliere expressie. Gebruik hierbij het standaard algoritme gebaseerd op "expression graphs".

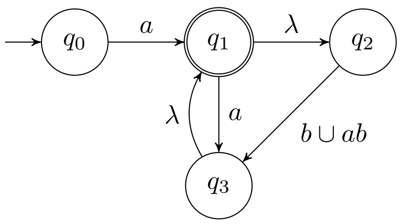
Antwoord op Opgave 1

a. *(8 punten)*

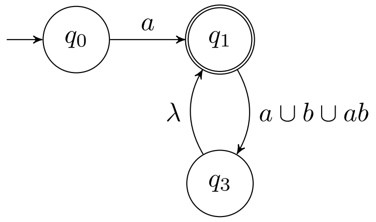


b. *(12 punten, 3 per correct uitgevoerde stap. Merk op dat de hier gekosten volgorde wel de simpelste, maar niet de enig mogelijke is!)*

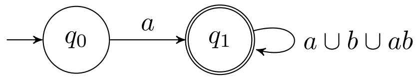
Na verwijderen van q_4:



Na verwijderen van q_2:



Na verwijderen van q_3:



De reguliere expressie wordt dan: $a(a \cup b \cup ab)^*$.

Opgave 2 **(20 punten)**

Gegeven de reguliere expressie $E = a^*(a \cup b)^*b$.

a. Beschrijf een stappenplan om een reguliere expressie om te zetten in een equivalente DFA met een minimaal aantal toestanden.

b. Voer deze stappen uit om E om te zetten in een minimale DFA. Geef alle tussenresultaten, en maak gebruik van de standaardalgoritmes.

Antwoord op Opgave 2

a. *(2 punten, alleen als het helemaal goed is; anders 0 punten.)* Het stappenplan:

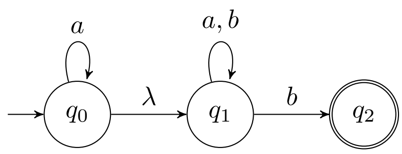
(i) RE naar NFA-λ via compositie;

(ii) Determinisatie van NFA-λ naar DFA;

(iii) Minimalisatie van DFA.

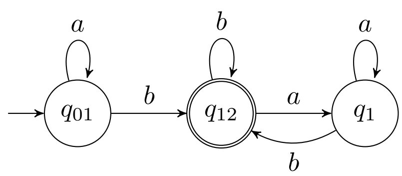
b. De uitvoering:

(i) *(6 punten.)* Als de compositiestappen geheel volgens het boek zijn uitgevoerd, resulteert dit in een vrij grote NFA met een heleboel overbodige λ's. Het vereenvourdigde resultaat is:



Geef de volle punten als dit antwoord goed is. Kijk alleen naar eventuele tussenresultaten als het antwoord fout is.

(ii) *(6 punten.)* Het resultaat na determinisatie (waarbij q_0q_1 afgekost is tot q_{01} etc.):

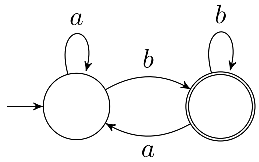


Geef de volle punten als dit antwoord goed is — inclusief het verband tussen de nieuwe en de oude toestanden. Kijk alleen naar eventuele tussenresultaten als het antwoord fout is.

(iii) *(6 punten.)* De tabel:

	q_{01}	q_{12}	q_1
q_{01}	$-$	0	$=$
q_{12}	$-$	$-$	0
q_1	$-$	$-$	$-$

Het resultaat na minimalisatie wordt:



Opgave 3 (15 punten)

Gegeven de volgende contextvrije grammatica G, met $V = \{S, P, Q\}$, $\Sigma = \{a, b\}$ en startsymbool S:

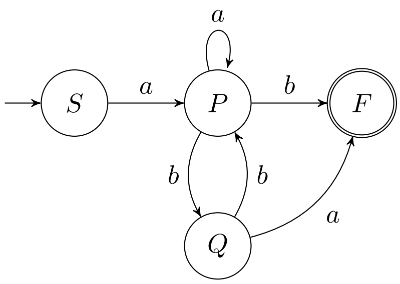
$$\begin{aligned} S &\to aP \\ P &\to aP \mid bQ \mid b \\ Q &\to bP \mid a \end{aligned}$$

a. G is regulier. Waarom?
b. Geef de bijbehorende eindige automaat
c. Geef in deze eindige automaat de berekening van het woorde $aabbb$
d. G is in Greibach normaalvorm. Waarom?
e. Geef de bijbehorende stapelautomaat
f. Geef in deze stapelautomaat de berekening van het woord $aabbb$

Antwoord op Opgave 3

a. *(2 punten)* In een reguliere grammatica moeten alle regels van één van de volgende vormen zijn: (i) $A \to \lambda$, (ii) $A \to aB$, of (iii) $A \to a$ (voor willekeurige $A, B \in V$ en $a \in \Sigma$). Voor G is dat inderdaad het geval.

b. *(3 punten)* De automaat:

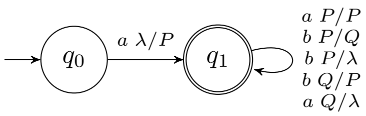


c. *(2 punten)* De berekening:

$$\begin{aligned} & [S, aabbb] \\ \vdash\ & [P, abbb] \\ \vdash\ & [P, bbb] \\ \vdash\ & [Q, bb] \\ \vdash\ & [P, b] \\ \vdash\ & [F, \lambda] \end{aligned}$$

d. *(2 punten)* In Greibach normaalvorm moeten alle regels van de vorm $A \to aA_1 \cdots A_n$ zijn (met $a \in \Sigma$ en $A_1, \ldots, A_n \in V \setminus \{S\}$). Voor G is dat inderdaad het geval (merk op dat $n = 0$ bij regels zoals $P \to b$).

e. *(3 punten)* De automaat:



f. *(3 punten)* De berekening:
$$\begin{aligned}
&[q_0, aabbb, \lambda]\\
\vdash\ &[q_1, abbb, P]\\
\vdash\ &[q_1, bbb, P]\\
\vdash\ &[q_1, bb, Q]\\
\vdash\ &[q_1, b, P]\\
\vdash\ &[q_1, \lambda, \lambda]
\end{aligned}$$

Opgave 4 **(20 punten)**

Gegeven de volgende contextvrije grammatica, met $V = \{E, F\}$, $\Sigma = \{1, +, (,)\}$ en startsymbool E:

$$\begin{aligned}
E &\to F \mid F{+}F\\
F &\to 1 \mid (E)
\end{aligned}$$

Deze grammatica genereert de taal van (zeer eenvoudige) *expressies*; bijvoorbeeld horen $(1 + 1) + 1$ en $1 + (1 + (1 + (1)))$ tot de taal. Als we voor de symbolen letters gebruiken, zoals gebruikelijk in het boek, bijvoorbeeld door de substitutie

$$1 \leftrightarrow e \qquad + \leftrightarrow p \qquad (\leftrightarrow a \qquad) \leftrightarrow b$$

dan verandert het alfabet in $\{a, b, e, p\}$, de grammatica in

$$\begin{aligned}
E &\to F \mid FpF\\
F &\to e \mid aEb
\end{aligned}$$

en de bovengenoemde voorbeeldwoorden worden $aepebpe$ en $epaepaepaebbb$. De opgaven hieronder gaan over deze, laatste, grammatica, die we G noemen.

a. Geef in G een linkerafleiding, een rechterafleiding en een afleidingsboom voor $aepebpe$.

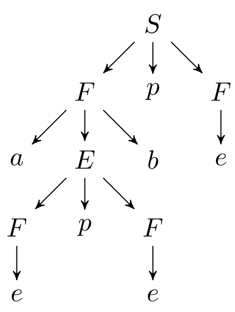
b. Zet G om naar Chomsky normaalvorm, volgens de in het boek beschreven methode. Licht de stappen die u neemt toe!

c. Zet voor de verkregen CNF-grammatica een CYK-matrix op die het woord $aepebpe$ ontleedt.

Antwoord op Opgave 4

a. linkerafleiding en rechterafleiding *(samen 2 punten, 0 als één van beide fout is)* en afleidingsboom (*3 punten*):

$$\begin{aligned}
S &\to FpF\\
&\to aEbpF\\
&\to aFpFbpF\\
&\to aepFbpF\\
&\to aepebpF\\
&\to aepebpe
\end{aligned}
\qquad\qquad
\begin{aligned}
S &\to FpF\\
&\to Fpe\\
&\to aEbpe\\
&\to aFpFbpe\\
&\to aFpebpe\\
&\to aepebpe
\end{aligned}$$



b. *(8 punten)* λ-regels zijn er niet, dus startrecursie hoeft eigenlijk ook niet verwijderd te worden. Hieronder gebeurt dit wél; trek niet meer dan 1 punt af als dit acterwege gebleven is. Na introductie van een nieuw startsymbool wordt de grammatica:

$$\begin{array}{l} S \to F \mid F{+}F \\ E \to F \mid F{+}F \\ F \to 1 \mid (E) \end{array}$$

De enige kettingregel is $E{\to}F$, dus krijgen we

$$\begin{array}{rcl} \mathit{CHAIN}(S) & = & S, F \\ \mathit{CHAIN}(E) & = & E, F \\ \mathit{CHAIN}(F) & = & F \end{array}$$

en wordt de grammatica

$$\begin{array}{l} S \to e \mid aEb \mid FpF \\ E \to e \mid aEb \mid FpF \\ F \to e \mid aEb \end{array}$$

Nu moeten we de regels nog in voldoende kleine stukjes hakken, door introductie van nieuwe variabelen. Dit kan op meerdere manieren; één manier is

$$\begin{array}{l} S \to e \mid AG \mid FH \\ E \to e \mid AG \mid FH \\ F \to e \mid AG \\ A \to a \\ G \to EB \\ B \to b \\ H \to PF \\ P \to p \end{array}$$

c. *(7 punten)* De CYK-matrix voor $aepebpe$, volgens bovenstaande grammatica:

$$\begin{array}{c|c|c|c|c|c|c} A & - & - & - & S,E,F & - & S,E \\ & S,E,F & - & E & G & - & - \\ & & P & H & - & - & - \\ & & & S,E,F & G & - & - \\ & & & & B & - & - \\ & & & & & P & H \\ & & & & & & S,E,F \end{array}$$

Opgave 5 **(10 punten)**

U wordt de nieuwe directeur van Softwarehouse BadIQ en moet saneren. Geef aan waarom de volgende twee projecten maar beter per direct gestopt kunnen worden:

a. Terminator: Om in een klap af te zijn van software die hangt, worden Java programma's automatisch omgezet naar Turingmachines. Die Turingmachines worden vervolgens automatisch geanalyseerd op terminatie.

b. Determinator: Om context-vrije grammatica's (CFG) te simuleren, wordt een willekeurige CFG eerst omgezet in een stapelautomaat (PDA), die vervolgens gedeterminiseerd wordt. Uit een deterministische PDA is immers eenvoudig code te genereren.

Antwoord op Opgave 5

a. *(5 punten; trek punten af als de uitleg niet dekkend is.)* Analyseren of een TM termineert is het halting probleem. Dat is onbeslisbaar, dus dit kan niet geautomatiseerd worden.

b. *(5 punten; trek punten af als de uitleg niet dekkend is.)* Er zijn context-vrije talen waarvoor geen deterministische PDA bestaat. (ofwel: de klasse van deterministische PDA-herkenbare talen is een strikt subset van de context-vrije talen).

Opgave 6 (15 punten)

Beschouw de volgende twee niet-reguliere talen:

$$\begin{array}{lcl} \mathcal{L}_1 & := & \{a^i b^j \mid i \leq j\} \\ \mathcal{L}_2 & := & \{w \in \{a,b\}^* \mid \#_a(w) \leq \#_b(w)\} \end{array}$$

waarbij $\#_a(w)$ het aantal voorkomens van a in w betekent.

a. Bewijs met behulp van de pompstelling dat $\mathcal{L}_1$ niet regulier is.
b. Bewijs met behulp van de geslotenheidseigenschappen van reguliere talen dat dan ook $\mathcal{L}_2$ niet regulier is.

Antwoord op Opgave 6

a. *(10 punten)* $\mathcal{L}_1 = \{a^i b^j \mid 0 \leq i \leq j\}$. Stel $\mathcal{L}_1$ is regulier. Laat $k > 0$ gegeven, kies $z = a^k b^k$, dan $|z| = 2k > k$, en $z \in \mathcal{L}_1$. Stel $z = uvw$ met $|uv| \leq k$ en $|v| \geq 1$. Laat $z' = uv^2w$. Volgens de pompstelling geldt dan $z' \in \mathcal{L}_1$.

   Echter, $u = a^p$, $v = a^q$ en $w = a^{k-p-q}b^k$ voor zekere p, q, met $q > 0$. dus $z' = uv^2w = a^{k+q}b^k$. Omdat $k + q > k$ geldt $z' \notin \mathcal{L}_1$. Tegenspraak. Dus $\mathcal{L}_1$ is niet regulier.

b. *(5 punten)* Stel $\mathcal{L}_2$ is regulier. Merk op dat $\mathcal{L}_1 = \mathcal{L}_2 \cap a^*b^*$. Reguliere talen zijn gesloten onder doorsnede, dus $\mathcal{L}_1$ is dan regulier. Tegenspraak. Dus $\mathcal{L}_2$ is niet regulier.