TENTAMEN

Basismodellen in de Informatica

vakcode: 211180
datum: 1 juli 2010
tijd: 8:45–12:15 uur

<u>VOORBEELDUITWERKING</u>

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 100 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.

Opgave 1 **(20 punten)**

We modelleren een stoplicht met het volgende alfabet: *g*(roen), *o*(ranje), *r*(ood) en *u*(it). De standaardcyclus van een stoplicht is een herhaald (groen, oranje, rood). Soms wordt deze onderbroken door oranjeknipperen. In dat geval wordt oranje vervangen door een herhaling van (oranje,uit), die altijd begint en eindigt met oranje.

a. Beschrijf dit gedrag met behulp van een reguliere expressie.

b. Zet deze expressie om in een NFA-λ. Gebruik hierbij het algoritme uit het boek (hierbij mag je loze λ-transities weglaten).

c. Determiniseer deze automaat via de subset-constructie.

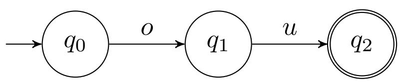
d. Minimaliseer de verkregen DFA. Ga hiervoor systematisch na welke toestanden equivalent zijn.

Laat bij de constructies in b–d de tussenstappen zien!

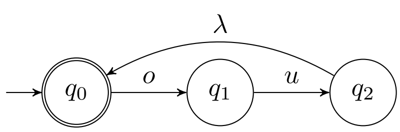
Antwoord op Opgave 1

a. *(5 punten)* Een korte oplossing: $(g.(o.u)^*.o.r)^*$

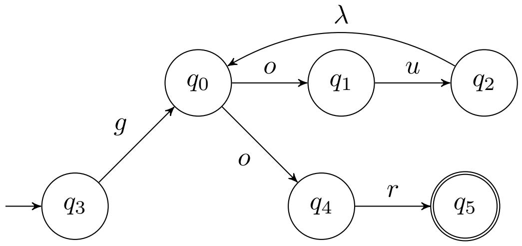
b. *(5 punten)* De constructie voor $o.u$:



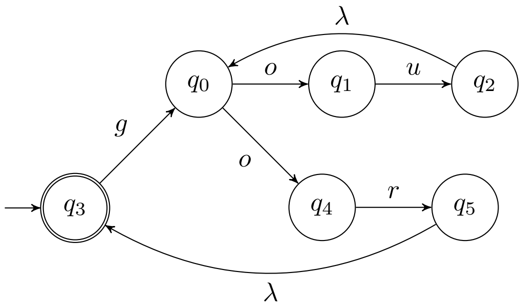
De constructie voor $(o.u)^*$:



De constructie voor $g.(o.u)^*.o.r$:



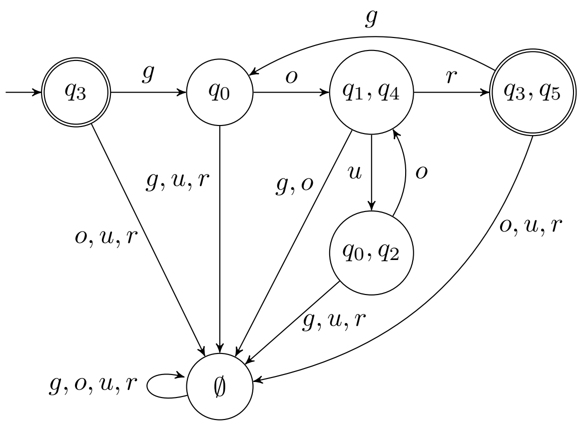
Tenslotte, de constructie voor $(g.(o.u)^*.o.r)^*$:



c. De transitietabel na λ-afsluiting:

δ	g	o	u	r
q_3	q_0			
q_0		q_1, q_4		
q_1			q_0, q_2	
q_2		q_1, q_4		
q_4				q_3, q_5
q_5	q_0			

De resulterende DFA:



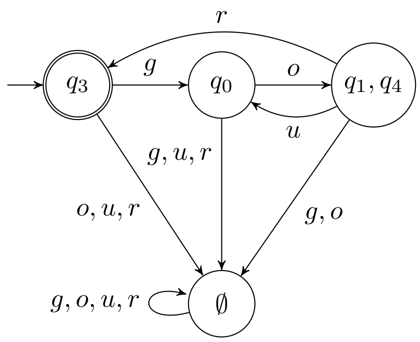
d. *(5 punten)* De minimalisatiematrix, na involling van accepterende toestanden (X = verschillend):

0	14	35	02	∅	
X	X		X	X	3
		X			0
		X			14
			X	X	35
					02

De rest is dan onmiddellijk ingevuld (geen look-ahead nodig):

0	14	35	02	$\emptyset$	
X	X	$=$	X	X	3
	X	X	$=$	X	0
		X	X	X	14
			X	X	35
				X	02

De geminimaliseerde automaat wordt dan:



Opgave 2 **(20 punten)**

Geef van de volgende stellingen aan of ze waar zijn (met een *kort* argument), dan wel onwaar (met een *klein* tegenvoorbeeld). Ongemotiveerde ja/nee-antwoorden worden niet goed gerekend.

a. Zij $M_1 = (Q, \Sigma, \delta, q_0, F)$ een willekeurige NFA, en zij $M_2 = (Q, \Sigma, \delta, q_0, Q - F)$. Dan geldt $L(M_1) = \overline{L(M_2)}$ (het laatste staat voor het complement van de taal, zie pag. 201 van het boek).

b. Voor zowel eindige automaten, stapelautomaten als Turingmachines geldt dat met de deterministische variant dezelfde klasse van talen gedefinieerd kan worden als met de non-deterministische variant.

c. Als een taal recursief opsombaar is, is zijn complement het ook. (Met andere woorden, de recursief opsombare talen zijn gesloten onder complement.)

Antwoord op Opgave 2

a. *(5 punten.)* Onwaar, bijvoorbeeld $M_1 = q_0 \rightarrow^a q_1, q_0 \rightarrow^a q_2$, met q_0 initieel en q_1 accepterend. Beiden accepteren het woord a..

b. *(5 punten.)* Onwaar, dit geldt niet voor pushdown automaten, met als standaardvoorbeeld uit het boek: $\{ww^R \mid w \in \{a, b\}*\}$.

c. *(10 punten.)* Onwaar. Een bewijs uit het ongerijmde: Neem aan dat de stelling waar is; dus dat er voor elke willekeurige recursief opsombare taal L, met accepterende Turingmachine M_1, een Turingmachine M_2 is die $\overline{L}$ accepteert. Construeer dan $M = M_1 \times M_2$ door M_1 en M_2 parallel te zetten, met gescheiden tapes, hetzelfde invoerwoord, en het product van de toestanden en transities. Een toestand in M is een succestoestand als de deeltoestand uit M_1 dat is.

Deze nieuwe machine stopt dus als M_1 stopt of M_2 stopt, en stopt in een succestoestand als M_1 dat doet. Aangezien elk woord óf in L óf in $\overline{L}$ zit, stopt M voor elk invoerwoord. De taal L is dus recursief. We hebben hiermee bewezen dat elke recursief opsombare taal recursief is, wat in tegenspraak is met het bestaan van onbeslisbare problemen zoals het halting problem. De stelling is dus onwaar.

Aanwijzing voor de correctie: Geef deelpunten voor antwoorden die "de goede kant opgaan", bijvoorbeeld minstens 5 punten als er een begin van een bewijs is en minstens 7 punten als het verband met recursieve talen wordt gelegd.

Opgave 3 (15 punten)

a. Laat zien dat de taal $\{(a^2b^3)^{2k} \mid k > 0\}$ regulier is.

b. Bewijs dat de taal $\{a^{4k}b^{6k} \mid k > 0\}$ niet regulier is.

Antwoord op Opgave 3

a. *(5 punten)* Dit kan bijvoorbeeld door de reguliere expressie: $(aabbbaabbb)(aabbbaabbb)*$.

b. *(10 punten)* Dit kan door de pompstelling. Stel dat hij (of zij) wel regulier is. Laat $k > 0$, kies $z := a^{4k}b^{6k}$. Merk op dat $z \in \mathcal{L}$ en $|z| > k$. Laat u, v, w willekeurig, zodat $|uv| < k$, $|v| > 0$ en $z = uvw$. Laat $p := |u|$ en $q := |v|$. Kies nu $i := 2$. Volgens de pompstelling voor reguliere talen geldt dan: $uvvw \in \mathcal{L}$. Echter $uvvw = a^p a^{2q} a^{4k-p-q} b^{6k}$. Dit zit alleen in $\mathcal{L}$ als voor zekere k' geldt: $6k = 6k'$ en $4k + q = 4k'$. Hieruit volgt $k = k'$, en dus $4k + q = 4k$, dus $q = 0$, tegenspraak. Dus deze taal is niet regulier.

Opgave 4 (20 punten)

Beschouw de volgende grammatica, met alfabet $\Sigma = \{\mathsf{hup}, \mathsf{holland}, \mathsf{hoera}, !\}$:

$$
\begin{aligned}
S &\rightarrow S\ \mathsf{hoera} \mid X \\
X &\rightarrow H\ X\ ! \mid \lambda \\
H &\rightarrow \mathsf{hup} \mid \mathsf{holland}
\end{aligned}
$$

a. Beschrijf (in verzamelingnotatie) welke taal deze grammatica genereert

b. Construeer stapsgewijs een Chomsky normaalvorm voor deze grammatica.

c. Construeer stapsgewijs een Greibach normaalvorm voor deze grammatica.

Antwoord op Opgave 4

a. *(4 punten)* De taal is $\{(\mathsf{hup} \cup \mathsf{holland})^i !^i \mathsf{hoera}^* \mid i \geq 0\}$. (Trek geen punten af voor een mengsel van eqruliere expressies en verzamelingen.)

b. *(7 punten)* Eerst moet de startrecursie verwijzerd worden. Met een nieuw startsymbool R:

$$
\begin{aligned}
R &\rightarrow S \\
S &\rightarrow S\ \mathsf{hoera} \mid X \\
X &\rightarrow H\ X\ ! \mid \lambda \\
H &\rightarrow \mathsf{hup} \mid \mathsf{holland}
\end{aligned}
$$

Nu moeten de λ-regels eruit. De λ-variabelen worden iteratief vastgesteld op $\{X\}$, daarna $\{S, X\}$, daarna $\{R, S, X\}$. Na inlinen en verwijderen van alle λ's behalve by het startsymbool:

$$\begin{aligned} R &\to S \mid \lambda \\ S &\to S\ \mathsf{hoera} \mid \mathsf{hoera} \mid X \\ X &\to H\,X\,! \mid H\,! \\ H &\to \mathsf{hup} \mid \mathsf{holland} \end{aligned}$$

Nu verwijderen we de kettingregels. De tabel met kettingregels is:

R	S, X
S	X
X	–
H	–

Dit resulteert in de grammatica

$$\begin{aligned} R &\to S\ \mathsf{hoera} \mid \mathsf{hoera} \mid H\,X\,! \mid H\,! \mid \lambda \\ S &\to S\ \mathsf{hoera} \mid \mathsf{hoera} \mid H\,X\,! \mid H\,! \\ X &\to H\,X\,! \mid H\,! \\ H &\to \mathsf{hup} \mid \mathsf{holland} \end{aligned}$$

Nu fatsoeneren we de regels door invoeren van extra variabelen:

$$\begin{aligned} R &\to S\ A \mid \mathsf{hoera} \mid H\,V \mid H\,U \mid \lambda \\ S &\to S\ A \mid \mathsf{hoera} \mid H\,V \mid H\,U \\ A &\to \mathsf{hoera} \\ V &\to X\,U \\ U &\to\ ! \\ X &\to H\,V \mid H\,U \\ H &\to \mathsf{hup} \mid \mathsf{holland} \end{aligned}$$

c. *(8 punten)* We kunnen uitgaan van elk van de laatste twee grammatica's hierboven; om het antwoorde klein te houden nemen we de op één na laatste. Het enige belangrijke dat gedaan moet worden is het verwijderen van de linksrecursie in S. Daarvoor een nieuwe variabele Z; de S-regel wordt dan

$$\begin{aligned} S &\to \mathsf{hoera}\ Z \mid H\,X\,!\,Z \mid H\,!\,Z \mid \mathsf{hoera} \mid H\,X\,! \mid H\,! \\ Z &\to \mathsf{hoera}\ Z \mid \mathsf{hoera} \end{aligned}$$

Als we $R < Z < S < X < H$ kiezen is de grammatica nu gestratificeerd. Met gebruik van inlinen, een extra variabele voor ! en het inzicht dat S en H nu onbereikbaar worden, krijgen we:

$$\begin{aligned} R &\to \mathsf{hoera}\ Z \mid \mathsf{hup}\ X\,U\,Z \mid \mathsf{holland}\ X\,U\,Z \mid \mathsf{hup}\ U\,Z \mid \mathsf{holland}\ U\,Z \mid \\ &\qquad \mathsf{hoera} \mid \mathsf{hup}\ X\,U \mid \mathsf{holland}\ X\,U \mid \mathsf{hup}\ U \mid \mathsf{holland}\ U \mid \lambda \\ Z &\to \mathsf{hoera}\ Z \mid \mathsf{hoera} \\ X &\to \mathsf{hup}\ X\,U \mid \mathsf{holland}\ X\,U \mid \mathsf{hup}\ U \mid \mathsf{holland}\ U \\ U &\to\ ! \end{aligned}$$

Opgave 5 (15 punten)

De Spaanse vlag bestaat uit twee horizontale rode banen, met daar tussenin een gele baan die twee keer zo breed is als elk van de rode. De gele baan bevat bovendien links het wapen van Spanje.

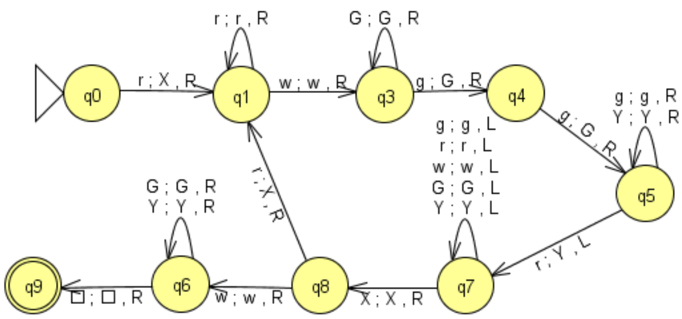
Definieer de taal van “spaanse-vlagwoorden”, bestaande uit een positief aantal keer r(ood), dan een enkele w(apen), dan twee keer zoveel g(eel) en dan weer het oorspronkelijke aantal r. Bijvoorbeeld zijn $rwggr$ en $rrwggggrr$ correcte spaanse-vlagwoorden.

a. Beschrijf de taal in verzamelingnotatie;
b. Geef een (single- of multi-tape) Turingmachine die de taal beslist, en geef de berekening voor $rrwggggrr$. Beargumenteer dat de machine de taal inderdaad beslist.
c. Geef een contextgevoelige grammatica die de taal genereert, en geef de afleiding voor $rrwggggrr$. Beargumenteer dat de grammatica inderdaad contextgevoelig is.

Antwoord op Opgave 5

a. *(3 punten)* De taal is $\{r^i\, w\, g^{2i}\, r^i \mid i > 0\}$.

b. *(6 punten. De hier gegeven oplossingen gaan uit van JFLAP; in een oplossing volgens het boek staat aan het begin van de tape altijd een blanco (B), die overgeslagen moet worden door een extra beweging naar rechts.)* Een oplossing met één tape:

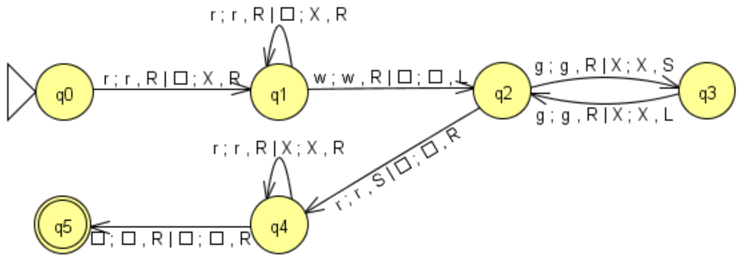


Deze grammatica beslist de taal omdat er in de hoofdlus telkens r’s in X’s of Y’s, dan wel g’s in G’s veranderd blijven worden, terwijl alle self-loops alleen laar links of rechts gaan en ophouden als er een blanco gelezen wordt. Deze herhalingen termineren altijd, dus de machine stopt altijd.

De gevraagde berekening (studenten hoeven dit niet in zoveel detail te laten zien, maar wel met een aantal tussentoestanden):

$$\begin{aligned} q_0rrwggggrr\Box &\vdash Xq_1rwggggrr\Box \vdash Xrq_1wggggrr\Box \vdash Xrwq_3ggggrr\Box \\ &\vdash XrwGq_4gggrr\Box \vdash XrwGGq_5ggrr\Box \vdash^* XrwGGggq_5rr\Box \\ &\vdash XrwGGgq_7gYr\Box \vdash^* q_7XrwGGggYr\Box \vdash Xq_8rwGGggYr\Box \\ &\vdash XXq_1wGGggYr\Box \vdash XXwq_3GGggYr\Box \vdash^* XXwGGq_3ggYr\Box \\ &\vdash^* XXwGGGGYq_5r\Box \vdash^* Xq_7XwGGGGYY\Box \vdash XXq_8wGGGGYY\Box \\ &\vdash XXwq_6GGGGYY\Box \vdash XXGGGGYYq_9\Box \end{aligned}$$

Een oplossing mat twee tapes:



Deze grammatica beslist de taal omdat de leeskop op tape 1 alleen maar naar rechts gaat onder het lezen van alfabetsymbolen. Omdat het invoerwoord eindig is stopt dit een keer.

De gevraagde berekening:

$$\begin{aligned}[q_0 rrwggggrr\Box, q_0\Box] &\vdash^* [rrq_0wggggrr\Box, XXq_0\Box] \vdash [rrwq_2ggggrr\Box, Xq_2X\Box] \\ &\vdash [rrwgq_3gggrr\Box, XXq_3\Box] \vdash [rrwggq_2ggrr\Box, q_2XX\Box] \\ &\vdash^* [rrwggggq_2rr\Box, q_2\Box XX\Box] \vdash [rrwggggq_4rr\Box, q_4XX\Box] \\ &\vdash^* [rrwggggrrq_4\Box, XXq_4\Box] \vdash^* [rrwggggr\Box q_5\Box, XX\Box q_5\Box]\end{aligned}$$

c. *(6 punten)* Een grammatica:

$$\begin{aligned} S &\to rwggr \mid rwMggr \\ M &\to XggY \mid XMggY \\ wX &\to rw \\ Yg &\to gY \\ Yr &\to rr \end{aligned}$$

Deze grammatica is contextgevoelig omdat hij niet-contraherend is: de rechterkant is bij alle regels langer dan de linkerkant.

De gevraagde afleiding:

$$S \Rightarrow rwMggr \Rightarrow rwXggYggr \Rightarrow rrwggYggr \Rightarrow rrwgggYgr \Rightarrow rrwggggYr \Rightarrow rrwggggrr \ .$$

Opgave 6 (10 punten)

a. Beschrijf hoe een schaakpartij als een woord van een formele taal opgevat kan worden; in het bijzonder, wat zijn de letters van het alfabet?

b. Waar in de Chomsky-hiërarchie bevindt zich de taal van alle schaakpartijen die door wit gewonnen worden? Beargumenteer je antwoord.

Antwoord op Opgave 6

a. *(2 punten)* De symbolen moeten een zet weergeven; dus ze moeten het beginveld en eindveld coderen (andere informatie, zoals de speler die aan zet is en welk stuk gezet wordt, is hieruit af te leiden). Meestal wordt hier een alfabet van de vorm $\Sigma = \{A, \ldots, H\}\{1, \ldots, 8\}\{A, \ldots, H\}\{1, \ldots, 8\}$ voor gebruikt.

b. *(8 punten)* Er zijn maar een eindig (hoewel buitengewoon groot) aantal toestanden nodig, nl. één voor elke mogelijke stand op het bord (met een paar subtiliteiten om de regels voor rocheren, en passant slaan en herhaling van zetten te verdisconteren). Een eindige automaat voldoet dus; oftewel, de taal is regulier.