TENTAMEN AUTOMATA THEORY
Donderdag 23 december 2021, 10.15 - 13.15 uur

Dit tentamen bestaat uit acht opgaven, waarbij steeds tussen [en] staat hoeveel punten er ongeveer mee te verdienen zijn. In totaal zijn er 100 punten te verdienen. Wanneer er bij een vraag om uitleg, motivatie of toelichting gevraagd wordt, is het belangrijk om die ook te geven.
Als we het in dit tentamen over een eindige automaat hebben (zonder verdere toevoeging), bedoelen we een deterministische eindige automaat zonder Λ-transities (wat elders *DFA* genoemd wordt).

---

1. [10 pt]

   (a) Teken een eindige automaat M, zó dat

   $$L(M) = \{a^i(ba)^j a^k \mid i,j,k \geq 1\}$$

   (b) Formeel is een eindige automaat een vijftal $(Q, \Sigma, q_0, A, \delta)$.
   Wat zijn voor je automaat M uit onderdeel (a) achtereenvolgens Q, Σ, q_0, A en $\delta^*(q_0, aaba)$? Geef concrete antwoorden voor M, dus geen algemene uitleg.

---

2. [9 pt] Laat

   $$L = \{x \in \{a,b\}^* \mid x = a^i(ba)^j a^k \text{ met } i,j,k \geq 0 \text{ en } x \text{ bevat evenveel voorkomens van de substring } ba \text{ als van de substring } aa\}$$

   Een string $x \in L$ moet dus aan alle vermelde eisen voldoen. Merk op dat voorkomens kunnen overlappen. Bijvoorbeeld $x = aaa$ bevat twee voorkomens van de substring aa.

   Inderdaad, deze taal L kennen we van de huiswerkopgaven. De eerste acht elementen in de canonieke volgorde van L zijn $\Lambda, a, baa, aaba, abaa, babaaa, aaababa, aababaa$.

   Toon aan dat de elementen van de oneindige verzameling $\{a^n \mid n \geq 0\}$ paarsgewijs *L-distinguishable* zijn. Ofwel, laat $m, n \geq 0$ willekeurig met $m \neq n$; toon aan, met behulp van de definitie van distinguishability, dat a^m en a^n L-distinguishable zijn.

---

3. [8 pt] Geef een reguliere expressie voor de taal

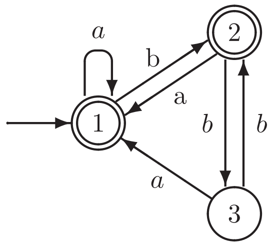
   $$L = \{x \in \{a,b\}^* \mid |x| \geq 2 \text{ en } x \text{ begint en eindigt met } aa \text{ of } ba\}$$

   Enkele elementen van L zijn bijvoorbeeld $aababa$ en $aaaaaaa$.

---

4. [15 pt] Tijdens de colleges hebben we twee (gerelateerde) constructies behandeld om op systematische wijze een reguliere expressie te construeren die correspondeert met een eindige automaat. De ene methode was de state-elimination methode van Brzozowski en McCluskey. De tweede methode bouwt reguliere expressies $r^k(i,j)$ op,[1] overeenkomend met de strings die je in de automaat voeren van toestand i naar toestand j, waarbij onderweg alleen toestanden 1 t/m k mogen worden gepasseerd. Hierbij nemen we aan dat de toestanden genummerd zijn als 1,2,.... Eerst worden alle expressies $r^0(i,j)$ bepaald, dan de expressies $r^1(i,j)$, dan de expressies $r^2(i,j)$, enzovoort.

   Het is de bedoeling om deze tweede constructie **gedeeltelijk** uit te voeren voor onderstaande eindige automaat M:



   Geef de volledige resulterende matrices r^0 en r^1, dus met waardes $r^0(i,j)$ en $r^1(i,j)$ voor alle i en j. Geef ook element $r^2(3,1)$. Hoewel we hiermee nog geen complete reguliere expressie hebben voor $L(M)$, is dit voldoende voor deze opgave.

   Je mag de reguliere expressies $r^1(i,j)$ die ontstaan vereenvoudigen tot equivalente expressies. De expressie $r^2(3,1)$ mag je niet vereenvoudigen (uitgaande van je reguliere expressies $r^1(i,j)$).

---

5. [16 pt] Laat opnieuw

$$\begin{aligned} L \;=\; \{x \in \{a,b\}^* \;\mid\; & x = a^i(ba)^j a^k \text{ met } i,j,k \geq 0 \\ & \text{en } x \text{ bevat evenveel voorkomens van de substring } ba \text{ als van de substring } aa\} \end{aligned}$$

   Een string $x \in L$ moet dus aan alle vermelde eisen voldoen. Merk op dat voorkomens kunnen overlappen. Bijvoorbeeld $x = aaa$ bevat twee voorkomens van de substring aa. Bij huiswerkopgave 3 werd gevraagd om een context-vrije grammatica voor L.

   (a) Laat G_1 de context-vrije grammatica zijn met startvariabele S en de volgende producties:

$$\begin{aligned} S &\rightarrow AB \mid aAB \mid \Lambda \\ A &\rightarrow aAba \mid \Lambda \\ B &\rightarrow baBa \mid \Lambda \end{aligned}$$

      i. Is $L(G_1) \subseteq L$? Zo nee, geef een element van $L(G_1)$ dat niet in L zit.

      ii. Is $L \subseteq L(G_1)$? Zo nee, geef een element van L dat niet in $L(G_1)$ zit.

[1] In het boek wordt deze expressies als $r(i,j,k)$ genoteerd.

(b) Laat G_2 de context-vrije grammatica zijn met startvariabele S en de volgende producties:

$$\begin{array}{rcl} S & \rightarrow & aA \ | \ B \ | \ aaAbabaBa \\ A & \rightarrow & aAba \ | \ \Lambda \\ B & \rightarrow & baBa \ | \ \Lambda \end{array}$$

i. Is $L(G_2) \subseteq L$? Zo nee, geef een element van $L(G_2)$ dat niet in L zit.

ii. Is $L \subseteq L(G_2)$? Zo nee, geef een element van L dat niet in $L(G_2)$ zit.

---

6. [17 pt] Laat

$$L = \{a^i b^j a^k \ \ | \ \ i, j, k \geq 0 \text{ en } i + k \leq 2j\}$$

(a) Geef de eerste zes elementen in de canonieke volgorde van L.

(b) Teken een stapelautomaat M, zó dat $L(M) = L$.

Deze stapelautomaat moet rechtstreeks gebaseerd zijn op eigenschappen van de taal. Hij moet dus niet het resultaat zijn van een standaard-constructie om b.v. een context-vrije grammatica om te zetten in een stapelautomaat.

Probeer ervoor te zorgen dat M deterministisch is. en geen Λ-transities bevat. Als dit niet lukt, kun je nog wel een deel van de punten verdienen.

Leg ook uit hoe M zijn toestanden en stapel gebruikt om precies de juiste taal te accepteren.

---

7. [12 pt]

(a) Laat $M = (Q, \Sigma, \Gamma, q_0, Z_0, A, \delta)$ een willekeurige stapelautomaat zijn.

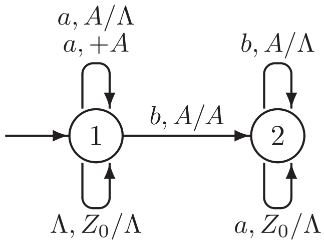
Wat betekent het wanneer we zeggen dat een string $x \in \Sigma^*$ door M geaccepteerd wordt *met lege stapel*? Je mag je antwoord formuleren in woorden en/of configuraties, als het maar duidelijk en volledig is.

Je verdient geen punten door te antwoorden dat x in de lege-stapel-taal $L_e(M)$ van M zit, want $L_e(M)$ is juist gedefinieerd als de verzameling van alle strings die door M met lege stapel worden geaccepteerd.

Als je het antwoord op dit onderdeel niet weet, dan kun je het 'kopen' van de docent. Wellicht kun je dan wel onderdeel (b) maken.

(b) i. Leg uit hoe je een willekeurige stapelautomaat $M = (Q, \Sigma, \Gamma, q_0, Z_0, A, \delta)$ kunt ombouwen naar een andere stapelautomaat M_1, zó dat $L(M_1) = L_e(M)$. Ofwel: M_1 accepteert 'op de normale manier' (met accepterende toestanden) de taal die M met lege stapel accepteert.
N.B.: Je hoeft niet aan te tonen dat de constructie die je beschrijft correct is. (Hij moet natuurlijk wel correct zijn.)

ii. Laat nu M de volgende PDA zijn (met initieel stapelsymbool Z_0):



Teken de stapelautomaat M_1 die het resultaat is van de constructie uit het vorige onderdeel, zodat $L(M_1) = L_e(M)$.

---

8. [13 pt] Het pomplemma voor context-vrije talen luidt als volgt:

> Stel L is een context-vrije taal.
> Dan is er een integer n, zó dat
> voor iedere $u \in L$ waarvoor $|u| \geq n$, u geschreven kan worden als
> $u = vwxyz$ voor bepaalde strings v, w, x, y en z waarvoor
> 1. $|wy| \geq 1$ (dwz $wy \neq \Lambda$).
> 2. $|wxy| \leq n$.
> 3. Voor elke $m \geq 0$ behoort de string vw^mxy^mz ook tot L.

Laat nu

$$L_1 = \{a^i b^{i+k} a^k \mid 0 \leq i < k\}$$

en laat n voor deze taal het getal uit het pomplemma zijn.

Geef bij elk van de onderstaande vier strings u_1, u_2, u_3, u_4 aan of hij geschikt is om een tegenspraak met het pomplemma te vinden. Geef bovendien voor elke string x_i die **niet** geschikt is, aan waarom hij niet geschikt is, bijvoorbeeld via een concrete opsplitsing $vwxyz$ van u_i die keurig aan het pomplemma voldoet.

$$\begin{aligned} u_1 &= aabbbbbaaa \\ u_2 &= a^{n-1}b^{2n-1}a^n \\ u_3 &= a^n b^{2n+n} a^{2n} \\ u_4 &= a^n b^{2n+1} a^{n+1} \end{aligned}$$

`einde tentamen`