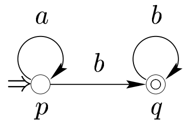
Formele Talen en Automaten 1 **5 januari 2000, 19.00 - 22.00**

Er zijn 5 opgaven die respectievelijk 15, 25, 20, 20 en 20 punten waard zijn. Voorzie de antwoorden van uitleg over de aanpak en van voldoende tussenresultaten. Geef indien van toepassing de nummers (uit het dictaat) van ieder resultaat dat wordt gebruikt.

1) Gegeven is de deterministische eindige automaat $\mathcal{A}$:



a. Construeer een met $\mathcal{A}$ equivalente eindige automaat in structurele normaalvorm (zonder λ-takken).

b. Bestaat er een deterministische eindige automaat in structurele normaalvorm die equivalent is met $\mathcal{A}$? Waarom (niet)?

c. Geef een MSOL($\{a,b\}$) formule voor de taal $L(\mathcal{A})$.

d. Geef een reguliere expressie voor de taal $\{a,b\}^* - L(\mathcal{A})$.

2) Voor iedere taal $L \subseteq \{a,b\}^*$ definiëren we
$sumir(L) = \{uvw \mid u,v,w \in \{a,b\}^*, uv \in L, w = \mathrm{mir}(v)\}$.

a. Bewijs dat de taal $K = sumir(\{a\}^*\{b\})$ niet representeerbaar is.

b. Laat zien dat K uit het vorige onderdeel wel context-vrij is.

c. Bewijs dat voor elke representeerbare taal R de taal $sumir(R)$ context-vrij is.

d. Geef een context-vrije taal L zodat $sumir(L)$ niet context-vrij is. (Zonder bewijs; wel een toelichting geven.)

3) Bekijk de context-vrije grammatica G met niet-terminalen S, A en B, terminalen a, b en c, startsymbool S, en met de volgende producties:

$S \to AB, \quad S \to Sb, \quad A \to AbBa, \quad A \to \lambda, \quad B \to bBc, \quad B \to bc.$

a. Geef een met G equivalente cf grammatica die λ-vrij is.

b. Geef een met G equivalente cf grammatica in Chomsky normaalvorm.

c. Geef een met G equivalente cf grammatica in Greibach normaalvorm.

4) H is de context-vrije grammatica met niet-terminalen S, A en B, terminalen a, b en c, startsymbool S, en met de volgende producties:

$S \to ABA, \quad A \to aAb, \quad A \to \lambda, \quad B \to cB, \quad B \to b.$

a. Toon aan dat de context-vrije grammatica G uit opgave 3 niet sterk LL(1) is.

b. Bereken voor grammatica H:
$first_1(C)$, $follow_1(C)$, $LA_1(C \to \gamma)$ voor elke niet-terminaal C en elke productie $C \to \gamma$.

c. Laat nu zien dat H sterk LL(1) is.

d. Geef aan hoe de LL(1)-ontleder voor H te werk gaat bij ontleding van het woord *cbaabb*. Doe dit door het geven van de rij van stapelinhouden en de bijbehorende producties (zoals in de tabel van voorbeeld 5.44).

5) $L = \{a^k b^l c^m \mid k, l, m \geq 0, k \neq l \text{ of } k \neq m\}$.
Ga na welke van de volgende uitspraken over L waar zijn.
Bewijs uw antwoord. (Hint: Kijk ook naar
$\overline{L} \cap \{a\}^*\{b\}^*\{c\}^* = \{a^k b^l c^m \mid k, l, m \geq 0, k = l = m\}$
waarbij $\overline{L}$ het complement van L is.)

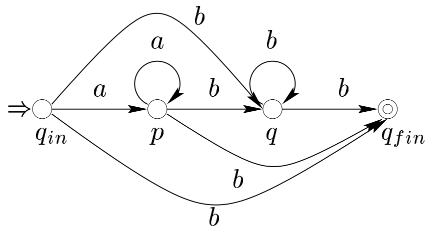
a. $L \in SA$

b. $L \in DSA_f$

c. $L \in DSA_e$

d. $L \in REP$

1) a. Een met $\mathcal{A}$ equivalente eindige automaat in structurele normaalvorm geconstrueerd als in het bewijs van Stelling 3.15:



b. Er bestaat geen deterministische eindige automaat in structurele normaalvorm die equivalent is met $\mathcal{A}$: zowel b als bb worden geaccepteerd door $\mathcal{A}$, dus $L(\mathcal{A})$ is niet prefix-vrij. Volgens Stelling 3.17 zal dus elke deterministische automaat die $L(\mathcal{A})$ accepteert minstens één eindtoestand moeten hebben met een uitgaande tak. Zo'n automaat zal dus nooit in SNF zijn.

c. Een MSOL($\{a, b\}$) formule voor de taal $L(\mathcal{A})$ direct uit $\mathcal{A}$ als in het bewijs van Stelling 3.52:

$\phi_{\mathcal{A}} = (\exists X_p)(\exists X_q)[(\exists i)\text{first}(i)$

$\wedge(\forall i)(i \in X_p \vee i \in X_q) \wedge (\forall i)\neg(i \in X_p \wedge i \in X_q)$

$\wedge(\forall i)(\text{first}(i) \rightarrow ((a(i) \wedge i \in X_p) \vee (b(i) \wedge i \in X_q))$

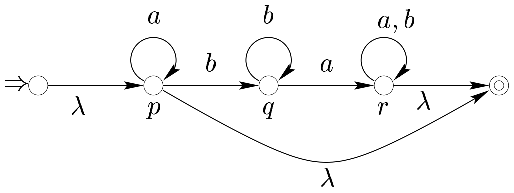
$\wedge(\forall i)(\forall j)(\text{next}(i, j) \rightarrow ((i \in X_p \wedge a(j) \wedge j \in X_p) \vee$

$(i \in X_p \wedge b(j) \wedge j \in X_q) \vee (i \in X_q \wedge b(j) \wedge j \in X_q)))$

$\wedge(\forall i)(\text{last(i)} \rightarrow i \in X_q)]$

Omdat $\mathcal{A}$ de taal $\{a\}^*\{b\}^+$ accepteert is ook de volgende formule goed: $\psi = (\exists i)(\forall j)((j < i \rightarrow a(j)) \wedge (j \geq i \rightarrow b(j)))$.

d. Maak $\mathcal{A}$ totaal en verwissel de rol van eindtoestanden en niet eindtoestanden. Geef vervolgens een equivalente eindige automaat in λSNF:



De parallelle takken van r naar r met labels a en b worden vervangen door één tak met label $a+b$. Achtereenvolgens de toestanden p, q en r verwijderen, leidt volgens de methode van Brzozowski tot de expressie: $\lambda \circ a^* \circ \lambda + \lambda \circ a^* \circ b \circ b^* \circ a \circ (a+b)^* \lambda$.
Zonder Brzozowski: Omdat $\mathcal{A}$ de taal $\{a\}^*\{b\}^+$ accepteert, bestaat $\{a,b\}^* - L(\mathcal{A})$ uit alle woorden die geen b bevatten of waarin na een b nog een a volgt. Dus een expressie voor het complement van $L(\mathcal{A})$ is: $a^* + a^* \circ b \circ b^* \circ a \circ (a+b)^*$.

2) a. $K = sumir(\{a\}^*\{b\}) = \{a\}^*\{b\} \cup \{a^n bba^m \mid n \geq m \geq 0\} \notin REP$.
Gebruik het pomplemma voor representeerbare talen (Stelling 3.18).
Laat $z = a^n bba^n$ waarbij n de constante van het pomplemma is.
Dan is $z = uvw$ met $|uv| \leq n$. Dus $v = a^k$ voor een $k \geq 1$ en als K aan het pomplemma voldoet dan is $a^{n-k}a^{i.k}bba^n \in K$ voor alle $i \geq 0$.
Pomp nu 'omlaag' ($i = 0$): $a^{n-k}bba^n \notin K$.
Hieruit concluderen we dat K niet aan het pomplemma voldoet.
Dus $K \notin REP$.

b. K wordt gegenereerd door de cf grammatica met axioma S en producties: $S \to aS|b|T, \quad T \to aTa|bb,$

c. Neem een rechts-lineaire grammatica voor R.
Voeg voor elke niet-terminaal A een nieuwe niet-terminaal $\overline{A}$ toe.
Voeg de volgende producties toe:
voor elke niet-terminaal A een productie $A \to \overline{A}$;
voor elke productie $A \to aB$ een nieuwe productie $\overline{A} \to a\overline{B}a$;
voor elke productie $A \to \lambda$ een nieuwe productie $\overline{A} \to \lambda$.
Deze grammatica begint zijn afleidingen zoals de oorspronkelijke rechts-lineaire. Nu wordt ofwel een woord uit R afgeleid (het gespiegelde suffix is λ) ofwel voor het laatste stuk van de afleiding overgegaan op niet-terminalen $\overline{A}$ die een suffix met zijn spiegelbeeld afleiden.

d. De taal $L = \{a^n ba^n \mid n \geq 0\}$ is context-vrij, terwijl $sumir(L) = \{a^n ba^m \mid n \leq m \leq 2n\} \cup \{a^n ba^{2n} ba^m \mid 0 \leq m \leq n\}$ niet context-vrij is.
Pas het pomplemma toe op $z = a^n ba^{2n} ba^n$.

3) a. Een met G equivalente cf grammatica die λ-vrij is:
$S \to AB|B|Sb, \quad A \to AbBa|bBa, \quad B \to bBc|bc.$

b. Chomsky normaalvorm: eerst moeten we een met G equivalente λ-vrije en ketenvrije grammatica hebben. We gaan verder met de grammatica geconstrueerd in het vorige onderdeel en maken die ketenvrij:
$S \to AB|bBc|bc|Sb, \quad A \to AbBa|bBa, \quad B \to bBc|bc.$
Vervolgens Chomsky:
$S \to AB|N_bX|N_bN_c|SN_b, \quad A \to AY|N_bZ, \quad B \to bX|N_bN_c,$
$X \to BN_c, \quad Y \to N_bZ, \quad Z \to BN_a,$
$N_a \to a, \quad N_b \to b, \quad N_c \to c.$

c. We gaan uit van de grammatica uit onderdeel a die al λ-vrij is.

stap 1 *(rec)* voor S: $S \to AB|B|ABX|BX$, $X \to bX|b$.

stap 2 *(sub)*: er zijn geen producties $A \to S\alpha$.
(rec) voor A: $A \to bBa|bBaY$, $Y \to bBaY|bBa$.
(sub) voor $S \to AB|ABX$:
. $S \to bBaB|bBaYB|B|bBaBX|bBaYBX|BX$.

stap 3 *(sub)*: er zijn geen producties $B \to S\alpha$ of $B \to A\alpha$.
(rec) voor B: er zijn geen producties $B \to B\alpha$.
(sub) voor $S \to B|BX$:
. $S \to bBaB|bBaYB|bBc|bc|bBaBX|bBaYBX|bBcX|bcX$.
(sub): er zijn geen producties $A \to B\alpha$.

X, Y *(sub)*: er zijn geen producties $X \to C\alpha$ of $Y \to C\alpha$.

Terminalen in orde maken levert ten slotte:
$N_a \to a$, $N_c \to c$, ($N_b \to b$ is niet nodig),
$S \to bBN_aB|bBN_aYB|$
. $bBN_c|bN_c|bBN_aBX|bBN_aYBX|bBN_cX|bN_cX,$
$A \to bBN_a|bBN_aY,$
$B \to bBN_c|bN_c,$
$X \to bX|b,$
$Y \to bBN_aY|bBN_a.$

4) a. De grammatica G uit opgave 3 bevat links-recursieve producties en is dus (Lemma 5.37) niet sterk LL(1).

b. Voor H geldt dat $first_1(C) = F_2(C)$ en $follow_1(C) = FL_2(C)$ zoals berekend in de volgende twee tabellen.

	S	A	B
F_0	$\emptyset$	a, λ	b, c
F_1	a, b, c	a, λ	b, c
F_2	a, b, c	a, λ	b, c

	S	A	B
FL_0	λ	b, c	a
FL_1	λ	b, c, λ	a, λ
FL_2	λ	b, c, λ	a, λ

$LA_1(C \to \gamma)$ voor elke niet-terminaal C en elke productie $C \to \gamma$:

$C \to \gamma$	$first_1(\gamma)$	$follow_1(C)$	$LA_1(C \to \gamma)$
$S \to ABA$	a, b, c	λ	a, b, c
$A \to aAb$	a	b, c, λ	a
$A \to \lambda$	λ	b, c, λ	b, c, λ
$B \to cB$	c	a, λ	c
$B \to b$	b	a, λ	b

c. H is sterk LL(1) omdat voor elke niet-terminaal C geldt dat voor elk tweetal verschillende producties $C \to \gamma_1$ en $C \to \gamma_2$ de verzamelingen $LA_1(C \to \gamma_1)$ en $LA_1(C \to \gamma_2)$ disjunct zijn (Lemma 5.35).

d. Ontleden van het woord *cbaabb*:

$\downarrow$	nog te lezen	stapel	productie
c	$baabb$	$S \perp$	$S \to ABA$
c	$baabb$	$ABA \perp$	$A \to \lambda$
c	$baabb$	$BA \perp$	$B \to cB$
c	$baabb$	$cBA \perp$	-
b	$aabb$	$BA \perp$	$B \to b$
b	$aabb$	$bA \perp$	-
a	abb	$A \perp$	$A \to aAb$
a	abb	$aAb \perp$	-
a	bb	$Ab \perp$	$A \to aAb$
a	bb	$aAbb \perp$	-
b	b	$Abb \perp$	$A \to \lambda$
b	b	$bb \perp$	-
b	λ	$b \perp$	-
λ	λ	$\perp$	-

5) a. $L \in SA$ is waar: L wordt gegenereerd door de context-vrije grammatica met axioma S en producties:
$S \to AX|YC$, $A \to aA|\lambda$, $B \to bB|\lambda$, $C \to cC|\lambda$,
$X \to bXc|bB|cC$, $Y \to aYb|aA|bB$.
Dus $L \in CF = SA$ (Stelling 5.13).

Een (niet-deterministische) stapelautomaat voor L zou eerst a's kunnen lezen en tellen (stapelen), vervolgens b's en dan c's lezen daarbij kiezend of het aantal b's danwel het aantal c's met het aantal a's wordt vergeleken (ontstapelen).

b. $L \in DSA_f$ is niet waar, want volgens Stelling 5.25 is DSA_f gesloten onder complement en doorsnede met reguliere talen. Dus als $L \in DSA_f$, dan ook $\overline{L} \cap \{a\}^*\{b\}^*\{c\}^* = \{a^k b^l c^m \mid k, l, m \geq 0, k = l = m\} \in DSA_f$, maar deze taal is niet eens context-vrij (pomplemma 4.42, zie ook gevolg 4.43) terwijl $DSA_f \subseteq SA = CF$ (Stelling 5.24).

c. $L \in DSA_e$ is niet waar, want $L \notin DSA_f$ en $DSA_e \subseteq DSA_f$ volgens Stelling 5.20.
Dat $L \notin DSA_e$ volgt ook uit het feit dat L niet prefix-vrij is (Lemma 5.18). Zo zijn bijv. zowel abc^3 als abc^2 element van L.

d. $L \in REP$ is niet waar, want $L \notin DSA_f$ en $REP \subseteq DSA_f$ volgens Stelling 5.19.
Dat $L \notin REP$ volgt ook, als bij onderdeel b, uit het feit dat REP gesloten is onder complement en doorsnijding met reguliere talen (Stelling 3.20) terwijl $\overline{L} \cap \{a\}^*\{b\}^*\{c\}^* \notin REP$.