

Opdracht 1

Topics on Parsing and Formal Languages - fall 2010

Rick van der Zwet
<hvdzwet@liacs.nl>

11 december 2010

Samenvatting

Dit schrijven zal uitwerkingen van opgaven behandelen uit het boek [JS2009] gebruikt bij het college. In deze opdracht zullen zeven opgaven (3,20,22,47,54,68,69) van hoofdstuk 3 behandeld worden.

1 Opgave 3.3

Als $L \subseteq \Sigma^*$ is regulier dan is de taal

$$2L := a_1, a_1, a_2, a_2, \dots, a_k, a_k : \text{elke } a_i \in \Sigma \text{ en } a_1 a_2 \dots a_k \in L \quad (1)$$

regulier. Zie dat er een ‘verdubbeling’ optreed van symbolen, dit gedrag is de modelleren in een *DFA*. Omdat L regulier is bestaat er een *DFA* $M = (Q, \Sigma, \delta, q_0, F)$ die L beschrijft. Construeer nu een nieuwe *DFA* P die de nieuwe taal $2L$ gaat beschrijven, neem hiervoor het alfabet (Σ), en de begin-toestand q_0 over van L . De toestanden (Q) worden verdubbeld. voor alle $q \in Q$: voeg q' toe aan Q . Neem ook de transities over van M , maar maak aanpassingen zodanig dat de nieuwe toestanden ook gelezen worden. Dus $\delta(q, x)$ wordt $\delta(q, q'), \delta(q', x)$. De acceptatie toestand F is gelijk aan de oude acceptatie toestand.

De nieuwe *DFA* P beschrijft $2L$ en dus is $2L$ regulier. \square

2 Opgave 3.20

Laat $\Sigma = \{0, 1\}$ zijn. Een voorbeeld van de taal $L \subseteq \Sigma^*$ voor welke geldt dat, de Myhill-Nerode[JS2009][pg. 77–81] gelijkheid relatie R_L de eigenschap heeft dat elk woord in Σ^* zijn eigen equivalentie klasse is. Dit wilt zeggen dat xR_Ly dan en slechts dan voor alle $z \in \Sigma^*$, is er een $xz \in L$ dan en slechts als $yz \in L$. Omdat hier gezocht wordt naar een taal waarbij de woorden allemaal equivalentie klassen op-zich zijn, zal alle elementen met elkaar vergeleken kunnen worden door middel van R_L en hier mogen geen positieve 'gevallen' uitkomen.

De (triviale) taal $L = 01, 10$ is een voorbeeld hiervan. Beiden woorden hebben hun eigen klasse. Als $x = 0$ en $y = 1$, bestaat er een $z = 1$, zodanig dat $xz \in L, yz \notin L$.

De taal $M = \{0^* : |M|_0 \text{ is een priemgetal}\}$ is een ander voorbeeld. Als we kijken naar aR_Lb kan er altijd een z gevormd worden zodanig dat $az \in M, bz \notin M$. Bijvoorbeeld $a = 0^3, b = 0^5, z = 0^4$. Deze eigenschap wordt afgedwongen door het 'feit' dat er geen relatie bestaat tussen de lengtes tussen opeenvolgende priemgetallen.¹

3 Opgave 3.22*

Om aan te tonen dat een $2DFA$ exponentieel meer expressief is dan een DFA voor bepaalde talen kijken we naar de volgende taal; Laat n een integer ≥ 1 en $F_n \subseteq \{0, 1, 2, 3, 4\}^*$ als volgende gedefinieerd:

$$F_n = \{3 0^{i_1} 1 0^{i_2} \dots 1 0^{i_n} 2^k 0^{i_k} 4 : 1 \leq k, j, i_j \leq n\} \quad (2)$$

Als eerste kan de F_n door een $2DFA$ van $O(n)$ toestanden geaccepteerd worden, welke aan te tonen is door eerst te kijken naar de eigenschappen van F_n hierbij is de zien dat de 0-reeks van de lengte i_k zowel voor als na de 2^k moet voorkomen. Waarbij 2^k aangeeft, waar deze 0-reeks gevonden kan worden bij de plek k . De $2DFA$ moet dus twee dingen doen. Het aantal keer nul op twee

¹Ik heb dit voor gemak aangenomen, de praktijk is echter dat is nog een groot wiskundig vraagstuk is.

plekken vergelijken. Waarbij 1 plek vast staat en de tweede plek variabele wordt gedefinieerd. De aanpak kan als volgt gezien worden:

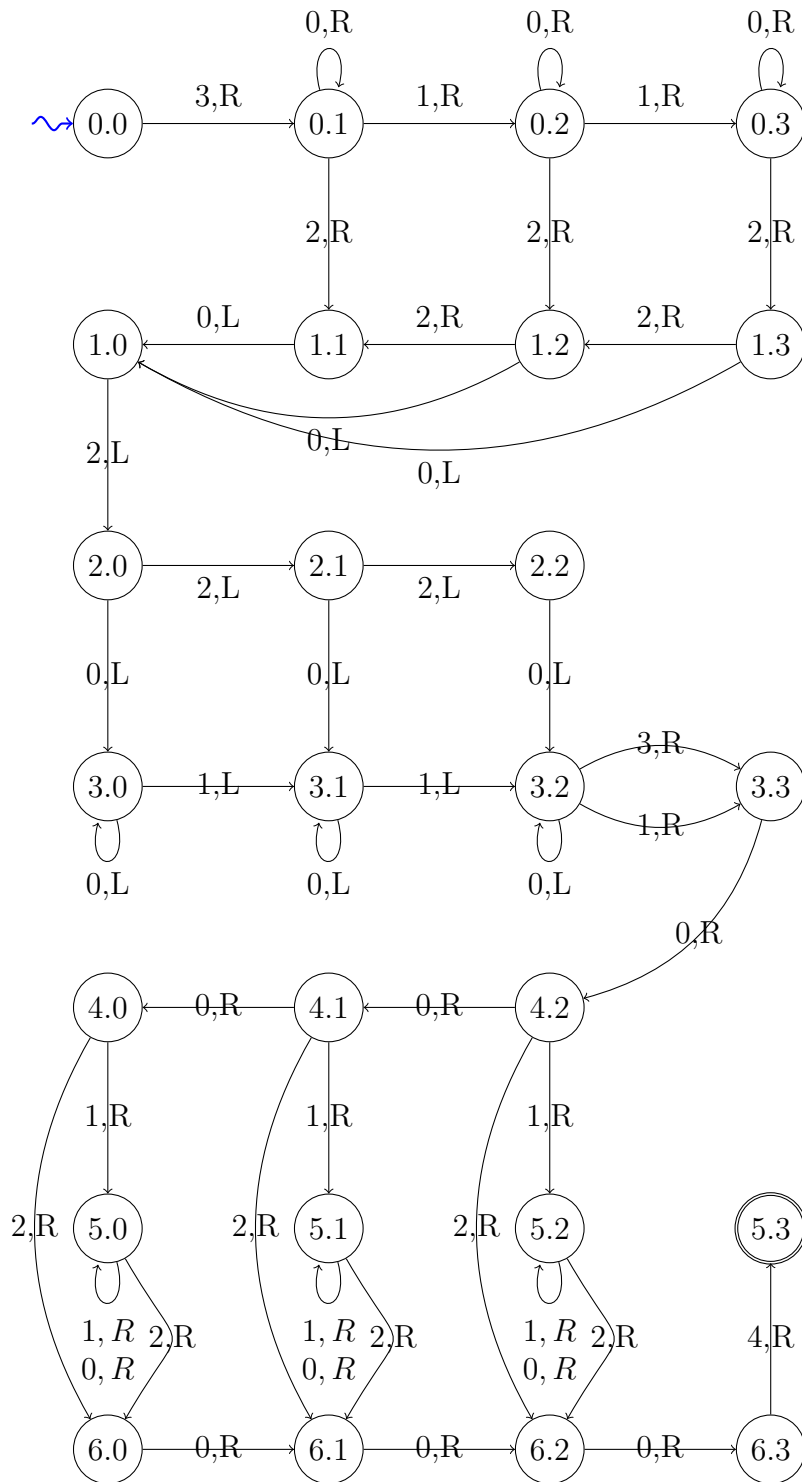
1. controleren of $|1| \leq 2$ de 0-reeks dmv 'knopen-strook' (knopen 0.X en 1.X).
2. $|2|$ tellen onthouden dmv 'knopen-strook' (knopen 2.X).
3. naar gewenste locatie lopen (knopen 3.X).
4. $|0|$ tellen onthouden dmv 'knopen-strook' (knopen 4.X).
5. naar begin 2de 0-reeks lopen (knopen 5.X).
6. 0-reeks vergelijken (knopen 6.X).

Voor deze aanpak zijn 3 'knopen-stroken' nodig, 3 'transitie-lagen' en 1 controle strook. In totaal dus in ongeveer $O(7n)$ knopen. Figuur 1 laat de *2DFA* zien.

Om te laten zien dat een *DFA* minstens n^n toestanden nodig heeft moet je gebruik maken van het feit dat een *DFA* niet terug kan lopen. Het zal dus een 'geheugen' moeten maken om het maximale 0 woord voor de 'splitsing' (2^k) te kunnen onthouden om zo te kijken of het overeen komt het 0 woord na de 'splitsing'. Omdat je voor elke F_i waarbij $1 \leq i \leq n$ minimaal n toestanden nodig hebt om te kunnen tellen en hiervan weer n unieke sets bestaan, heb je dus minimaal n^n toestanden nodig.

Dit vertaald zich in een Myhill-Nerode relatie om het aantal toestanden te kunnen bepalen. F_p en F_q waarbij $p < q, 1 \leq p, q \leq n$ is onderscheidbaar, door voor $z = 2*0*4$ te nemen, zodanig dat het woord door F_q geaccepteerd wordt, maar niet door F_p , door $|z|_2 \geq p$ te kiezen.

Voor elke $F_i : 1 \leq i \leq n$ geldt dat deze 'intern' ook i^2 unieke toestanden geeft. We splitsen de woorden bij 2, zodat we over een 'prefix' en 'suffix' kunnen praten. Een 'prefix' past maar bij 'e' en 'suffix'. De 'suffixes' die gegenereerd worden van de vorm $2*0*4$, kan i^2 vormen aannemen (bijvoorbeeld $i = 2204, 2004, 22004, 2204$). Deze unieke suffixes zorgen ervoor dat de Myhill-Nerode toepassing een substring z kan vinden welke een van de unieke suffixes bevat, waardoor het 'prefix' niet samenkomt met een 'suffix'.



Figuur 1: Opdracht 2.22a uitgewerkt voor $1 \leq n \leq 3$. Niet geldige stappen zijn niet genoemd en moet voor compleetheid naar een 'putje' knoop gestuurd worden. Uitbreiden kan door het expanden van 'triviale' knopen.

4 Opgave 3.47

Om te bewijzen dat ‘de klasse van talen geaccepteerd door *DFA*’ \Rightarrow ‘de klasse van talen gespecificeerd door de reguliere expressies’[JS2009][Theorem 1.4.2, (b) \Rightarrow (a)] aan de hand van het volgende voorbeeld: Laat $M = (Q, \Sigma, \delta, q_1, F)$ een *DFA* zijn, waarbij $Q = \{q_1, q_2, \dots, q_n\}$. Definieer nu $R_{i,j,k}$ als de taal van alle woorden die van toestand i naar toestand j gaan zonder door toestanden te gaan die hoger dan k genummerd zijn.

$R_{i,j,k}$ accepteert als $q_i \in q_1, q_j \in F$. Om $R_{i,j,k+1}$ te verkrijgen, nemen we eerst de toestanden van $R_{i,j,k}$, deze verenigen we dan met alle toestanden die we vanuit $R_{i,j,k}$ kunnen bereiken met een transitie $\delta(a, k+1) : i \leq a \leq j$, die in een nieuwe toestand uitkomen, de lussen hebben geen meerwaarde voor ons doel. De begin toestand is te markeren als $R_{i,j,0} = q_1 : i = j = 1$.

5 Opgave 3.54

Laat $\Sigma = \{1, 2, 3, \dots, n\}$ zijn en definieer:

$$L_n = \{w \in \Sigma^* : |w|_i = 1 \text{ voor alle } i \in \Sigma\} \quad (3)$$

Dit zijn dus de woorden waarbij alle elementen van Σ precies één keer voor komen. Bijvoorbeeld $L_3 = \{123, 132, 213, 231, 312, 321\}$.

a) Om te laten zijn dat er minimaal een reguliere expressie van lengte 2^{n-1} nodig is om L_n te specificeren, passen we de *Myhill-Nerode* stelling toe. Neem $x, y \in \Sigma^*$. Als geldt $|x| = |y|$ en de elementen zijn gelijk, maar enkel hun volgorde verschilt (bijvoorbeeld 123, 132) dan zitten ze in dezelfde klasse. Voor xR_iy als de lengte verschilt, dan is er een z te vinden zodanig dat y wel geaccepteerd wordt en x niet (bijvoorbeeld voor 34, 345, is dit 12), vanwege $|x| + a \neq |y| + a = n$ desals $|x| \neq |y|$. Als de ‘inhoud’ verschilt maar de lengte gelijk is dan $z = \Sigma - \{y\}$, omdat $\Sigma \setminus \{x\} \neq \Sigma \setminus \{y\}$.

b) Echter voor $\overline{L_n}$ is wel een reguliere expressie te vinden en wel in de grootte $O(n^2)$. $\bigcup_{a=0..n} \bigcup_{i \in \Sigma_a} \{x^*ix^*ix^*\}$ waarbij $x \in (\Sigma_a \setminus \{i\})^*$.

6 Opgave 3.68

Voor een woord $w \in \Sigma^*$, is een $\text{palc}(w)$ de korte palindroom x zodanig dat w een prefix is van x . en $\text{palc}(L) = \bigcup_{w \in L} \{\text{palc}(w)\}$.

a) Om te laten zien dat $\text{palc}(w) = wt^{-1}w^R$, waarbij wt^{-1} het woord is w waar het suffix t eraf gehaald is en t de langste palindroom suffix van w is, moet er gebruik gemaakt worden van een tegenstelling. Neem aan dat w een palindroom aan het einde zal bevatten uu en dus van de vorm is ruu dan ziet $\text{palc}(w)$ er als volgt uit $ruuuur$, dit is niet de korte palindroom (rr) , welke $uuuu$ er nog uit x weggehaald had kunnen worden. Een willekeurige suffix van w kan dus geen palindroom vormen die nog ‘weggesneden’ kan worden, omdat hij anders al in den beginne een palindroom had moeten zijn. t moet wel de langste palindroom suffix van w zijn, omdat anders een ‘dubbele’ palindroom problemen op gaat leveren. Neem bijvoorbeeld w van de vorm is $ruuyy$ als je van de palindroom yy verwijderd, zal $\text{palc}(w) = ruuuur$. Dit is niet de kortste palindroom (rr) .

b) Neem $L = \{vxyz : u, x, z \in 1^* \text{ en } v, u \in 0^*\}$ en $\Sigma = \{0, 1\}$. De woorden die $\text{palc}(L)$ moeten bevatten zijn alle woorden van L minus de woorden $|x|_1 = \text{even}, |v|_0 = |y|_0, |u|_0 = |z|_0$. Die set komt niet door het pumping lemma en is dus $\text{palc}(L)$ is niet regulier.

c) Neem $\text{palc}^{-1}(L)$, dan is dit $\{x \in L, u \in \Sigma^* : xuu\}$, welke alle woorden zijn met hun prefix in L en hun suffix is een palindroom. Het genereren van palindromen met een reguliere taal is niet mogelijk, dus $\text{palc}^{-1}(L)$ is niet regulier.

7 Opgave 3.69

Laat $x_1, x_2, \dots, x_k \in \Sigma^*$. Om te laten zien dat $\Sigma^* - x_1^*x_2^* \cdots x_k^*$ eindig is dan en slechts als $|\Sigma| = 1$ en $\gcd(|x_1|, |x_2|, \dots, |x_k|) = 1^2$, moeten beiden kanten op bekeken worden.

\Rightarrow : Door middel van een tegenstelling kunnen we bewijzen dat $|\Sigma| = 1$ moet zijn. Neem $|\Sigma| > 1$ en $x_1, x_2, \dots, x_n = 0$ dan is deze oneindig, en wel in de

²Ook paarsgewijs relatief priem genoemd, zie http://nl.wikipedia.org/wiki/Paarsgewijs_relatief_priem

vorm $(\Sigma \setminus 0)^*$. Door een tweede tegenstelling tonen we aan dat de gcd één moet zijn. Als $|\Sigma| = 1$ en $g = \gcd(|x_1|, |x_2|, \dots, |x_n|) \neq 1$ dan komen er enkel woorden voor van lengte $g^n : n = 1..n$. Hierdoor zullen de woorden met lengte $(g - 1)^*$ een oneindige reeks vormen.

\Leftarrow : Doordat $|\Sigma| = 1$ is de lengte van de string indentiek aan zijn formaat. Met een $\gcd(|x_1|, |x_2|, \dots, |x_n|) \neq 1$ kunnen we alle bijna alle lengtes genereren door de eigenschap dat de nummers (relatief) priem zijn.³

Referenties

[JS2009] Jeffrey Shallit, *A second course in formal languages and automata theory*, Cambridge University Press, 2009.

³Dit is een grote gok, aangenomen dat er een Stelling bestaat, welke stelt dat je uit 2 priemgetallen vanaf een bepaald punt alle getallen kan genereren.