

# Opdracht 2

Topics on Parsing and Formal Languages - fall 2010

Rick van der Zwet  
<hvdzwet@liacs.nl>

26 november 2010

## Samenvatting

Dit schrijven zal uitwerkingen van opgaven behandelen uit het boek [JS2009] gebruikt bij het college. In deze opdracht zullen zeven opgaven (9, 18, 24, 26, 32, 37, 41) van hoofdstuk 4 behandeld worden. De opgaven zijn willekeurig gekozen met behulp van een kans generator, het kan dus zijn dat niet alle onderwerpen van hoofdstuk 4 behandeld worden.

## 1 Opgave 4.9

Laat  $L \subseteq \Sigma^*$  de taal zijn en definieer  $\sigma(L) = \{x \in \Sigma^* : xy \in L \text{ voor alle } y \in \Sigma^*\}$ . Als  $L$  context-vrij is  $\sigma(L)$  ook context-vrij. Omdat  $y = \Sigma^*$  is dit een reguliere taal.  $xy \in L$  is vereniging van  $xy$  met  $L$ , welke onder context-vrije talen gesloten is. Hierdoor zal  $\sigma(L)$  (een subset) ook weer context-vrij zijn.

## 2 Opgave 4.18

Als  $L$  een context-vrije taal is en  $R$  een reguliere taal dan de rechter quotient  $L/R = \{x \in \Sigma^* : \exists y \in R \text{ zodat } xy \in L\}$  is ook een context-vrije taal.  $xy$

zijn de woorden in  $L$  waarbij geldt dat de suffix  $y$  regulier is. Stel dat  $x$  geen context-vrije taal was. Dan zou  $xy$  ook geen context-vrije taal zijn. Wat niet kan welke  $L$  context-vrij is en in die hoedanigheid ook de subsets van  $L$ .

### 3 Opgave 4.24

Als  $s^{[-1]}(L) := \{x : s(x) \cap L \neq \emptyset\}$  en  $s$  verbind letters met reguliere talen en  $L$  is context-vrij. Dan is  $s^{[-1]}(L)$  niet altijd CFL,  $L$  een verzameling kan zijn van alle talen van de letters waarbij geldt dat de lengte van de letters even moet zijn, welke geen context-vrije verzameling is.

Hetzelfde probleem doet voor als  $s$  letters aan context-vrije talen verbindt.

### 4 Opgave 4.26

Gegeven een PDA  $M$  kan je altijd een  $M'$  maken welke de eigenschap heeft dat geen enkele stap het huidige symbool op de stapel vervangt (geen transities van de vorm  $(p, \gamma Y) \in \delta(q, a, X)$  waarbij  $X \neq Y$ ). Dus dat alle stappen, of een symbool van de stapel afhalen, of een string van symbolen op de stapel zetten. Dit is te doen door de huidige 'substitutie' transities te vervangen door een tweetal (gekoppelde) transities. Waarbij de eerste het symbool eraf haalt en de tweede de gewenste symbolen er weer op zet.

De bovengenoemde methode zorgt voor maar één mogelijk pad. Er zal de de DPDA structuur (indien present) dus niet aantasten.

### 5 Opgave 4.32

Als  $\psi$  een Parikh map is, dan zijn er lange Engelse woorden over een bepaald alfabet voor specifieke eigenschappen. Voorbeelden zijn te vinden in de onderstaande tabel.

eigenschap	woorden	alfabet
$\psi(w)$ alle items gelijk aan 1	dutchwomen	$\{d, u, t, c, h, w, o, m, e, n\}$
$\psi(w)$ alle items gelijk aan 2	tomtom	$\{t, o, m\}$
$\psi(w)$ alle items gelijk aan 3	sestettes	$\{s, e, t\}$
$\psi(w)$ alle items $\geq 2$	unprosperousness	$\{e, n, o, p, r, s, u\}$
$\psi(w)$ (1, 2, 2, 3, 3, 3)	chinchinchee	$\{r, i, n, c, h, e\}$

## 6 Opgave 4.37

$\text{perm}(x)$  is de set van alle permutaties van de letters van  $x$ . Voor talen is  $\text{perm}(L) := \bigcup_{x \in L} \text{perm}(x)$ .  $\text{perm}(L)$  is niet context-vrij als  $L$  de reguliere taal is  $\{x \in a^*b^*c^* : |x|_a, |x|_b, |x|_c \text{ even zijn}\}$ .

[Dit is handen wapperen, ik zou graag wel willen weten hoe dit bewezen kan worden] Als  $L$  echter een reguliere taal over een alfabet van twee symbolen is, dan zal  $\text{perm}(L)$  context-vrij zijn. Welke de reguliere taal, maximaal één complexe relatie kan aangaan en wel namelijk tussen de twee symbolen in. Alle andere gevallen (enkel relatie één symbool) is er maar een enkelvoudige afhankelijk gecodeerd. Pak bijvoorbeeld de ingewikkelde  $\{x \in a^*b^* : |x|_a, |x|_b \text{ even zijn}\}$ . De permutatie van bijvoorbeeld  $aabb$  zullen zijn  $aabb, abba, abab, baba, baab, bbaa$ , waarbij in het context-vrije domain een analoge taal ontstaan, waarbij het aantal a ‘even’ moet zijn om te accepteren en het aantal b gelijk aan het aantal ‘a’.

## 7 Opgave 4.41

Om te bewijzen dat, de taal van alle woorden over de  $\{0, 1\}$  welke geen prefixen zijn van een Thue-Morse woord context-vrij is, moet er gekeken worden naar het meer generieke geval. Thue-Morse woorden zijn overlap-vrij, een context-vrije taal alle woorden opsomt welke een overlap bevatten is dus voldoende. Woorden met een overlap bevatten de serie  $cxcxc$  waarbij  $c$  een letter is en  $x$  een woord, de overlap is dan  $cxc$ . De taal van de woorden ziet er dus als volgt uit:  $\{0, 1\}^*1x1x\{0, 1\}^*$  of  $\{0, 1\}^*0x0x\{0, 1\}^*$  waarbij  $x := \{0, 1\}\{0, 1\}^*$ . Beiden delen zijn met een (niet deterministische) gramatica te maken, door expansie te doen vanuit ‘het midden’. Waarbij te denken is aan een soortgelijke setup (a,b ipv 0,1):

S  $\rightarrow$  MJ  
J  $\rightarrow$  aJ|bJ|a|b  
M  $\rightarrow$  a0a|bPb  
0  $\rightarrow$  a0a|b0b|a  
P  $\rightarrow$  aPa|bPb|b

## Referenties

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