U Vic CSC320 W2020 Homework 2 Solutions

1. (a) (5 marks) $S \to \epsilon \mid 0 \mid 1 \mid 0S0 \mid 1S1$

   (b) (5 marks) $S\epsilon \mid 0S1 \mid 0S1 \mid SS$

   (c) (5 marks) $S \to \epsilon \mid 0S1$

2. (10 marks)

$$\begin{aligned} S &\to AX \mid YC \\ A &\to \epsilon \mid aA \\ C &\to \epsilon \mid cC \\ X &\to \epsilon \mid bXc \\ Y &\to \epsilon \mid aYb \end{aligned}$$

   (10 marks) This is ambiguous. E.g., there are two leftmost derivations for ϵ: $S \Rightarrow AX \Rightarrow X \Rightarrow \epsilon$ and $S \Rightarrow YC \Rightarrow C \Rightarrow \epsilon$

3. (a) To start we just introduce a new start symbol A_0 (2 marks)

$$\begin{aligned} A_0 &\to A \mid \epsilon \\ A &\to BAB \mid B \mid \epsilon \\ B &\to \mathtt{00} \mid \epsilon \end{aligned}$$

   (b) Now eliminate ϵ-productions (other than for A_0) (2 marks)

$$\begin{aligned} A_0 &\to A \mid \epsilon \\ A &\to BAB \mid BB \mid AB \mid BA \mid A \mid B \\ B &\to \mathtt{00} \end{aligned}$$

   (c) Now eliminate unit productions (2 marks)

$$\begin{aligned} A_0 &\to BAB \mid BB \mid AB \mid BA \mid \mathtt{00} \mid \epsilon \\ A &\to BAB \mid BB \mid AB \mid BA \mid \mathtt{00} \\ B &\to \mathtt{00} \end{aligned}$$

   (d) Now we abbreviate any RHS with more than 2 symbols (2 marks)

$$\begin{aligned} A_0 &\to BC \mid BB \mid AB \mid BA \mid \mathtt{00} \mid \epsilon \\ A &\to BC \mid BB \mid AB \mid BA \mid \mathtt{00} \\ B &\to \mathtt{00} \\ C &\to AB \end{aligned}$$

   (e) Replace terminals that appear in RHS with more than one symbol by a variable (2 marks)

$$\begin{aligned} A_0 &\to BC \mid BB \mid AB \mid BA \mid ZZ \mid \epsilon \\ A &\to BC \mid BB \mid AB \mid BA \mid ZZ \\ B &\to ZZ \\ C &\to AB \\ Z &\to \mathtt{0} \end{aligned}$$

   (Full 10 marks for a grammar in this final form, even without all the intermediate steps.)

4. (15 marks) Using this version of the CNF

$$
\begin{aligned}
E &\to EA \mid EB \mid LD \mid \mathbf{id} \mid \mathbf{num} \\
A &\to ME \\
B &\to PE \\
D &\to ER \\
M &\to * \\
P &\to + \\
L &\to (\\
R &\to)
\end{aligned}
$$

$w = w_1w_2w_3w_4w_5w_6w_7 = (\mathbf{id} + \mathbf{num}) * \mathbf{num}$. Entry (i, j) of the table below contains all variables which derive $w_i \dots w_j$.

	1	2	3	4	5	6	7
1	L	–	–	–	E	–	E
2		E	–	E	D	–	–
3			P	B	–	–	–
4				E	D	–	–
5					R	–	–
6						M	A
7							E

5. The answer is "no" (2 marks).

   Consider the following grammar (others are possible): $S \to SS|a$. This is in CNF, but there are two leftmost derivations for aaa:

   $S \Rightarrow SS \Rightarrow aS \Rightarrow aSS \Rightarrow aaS \Rightarrow aaa$ and $S \Rightarrow SS \Rightarrow SSS \Rightarrow aSS \Rightarrow aaS \Rightarrow aaa$

   OR You could also just note that there are inherently ambiguous grammars, so converting to CNF won't help (3 marks for either way.)