Sample solutions for the examination of Finite automata and formal languages (DIT321/DIT322/TMV027/TMV028) from 2020-03-19

Nils Anders Danielsson

1. (a) Yes, the right-hand side of every production contains either a single terminal or exactly two nonterminals.

   (b) The CYK table:

$$
\begin{array}{cccc}
\{ S, A \} & & & \\
\{ S, A \} & \emptyset & & \\
\{ S, A \} & \emptyset & \emptyset & \\
\{ A \} & \{ B \} & \{ B \} & \{ B \} \\
\hline
a & b & b & b
\end{array}
$$

   (c) We have $abbb \in L(G, N)$ if and only if N is in the "top-most" cell of the CYK table for $abbb$. Thus we have $abbb \in L(G, S)$ and $abbb \in L(G, A)$, but not $abbb \in L(G, B)$.

2. The Turing machine is $(Q, \{ 0, 1, 2, 3 \}, \Gamma, \delta, \mathit{zero}, \sqcup, \{ \mathit{accept} \})$, where Q, Γ and δ are defined in the following way:

$$
\begin{array}{l}
Q = \{ \mathit{zero}, \mathit{one\text{-}or\text{-}three}, \mathit{two}, \mathit{blank}, \mathit{accept} \} \\
\Gamma = \{ 0, 1, 2, 3, \sqcup \} \\
\\
\delta \in Q \times \Gamma \rightharpoonup Q \times \Gamma \times \{ \mathsf{L}, \mathsf{R} \} \\
\delta(\mathit{zero}, 0) \qquad\qquad\;\; = (\mathit{one\text{-}or\text{-}three}, \sqcup, \mathsf{R}) \\
\delta(\mathit{one\text{-}or\text{-}three}, 1) = (\mathit{two}, \sqcup, \mathsf{R}) \\
\delta(\mathit{one\text{-}or\text{-}three}, 3) = (\mathit{blank}, \sqcup, \mathsf{R}) \\
\delta(\mathit{two}, 2) \qquad\qquad\;\;\; = (\mathit{one\text{-}or\text{-}three}, \sqcup, \mathsf{R}) \\
\delta(\mathit{blank}, \sqcup) \qquad\qquad = (\mathit{accept}, \sqcup, \mathsf{R})
\end{array}
$$

   The machine always moves to the right. It accepts if and only if it finds a 0, zero or more repetitions of 12, and a 3, followed by a blank.

3. (a) The ε-NFA A corresponds to the following system of equations between languages, where e_0 corresponds to the start state s_0:

$$\begin{aligned} e_0 &= e_1 \\ e_1 &= a(e_1 + e_2) + be_0 \\ e_2 &= \varepsilon + b(e_0 + e_1) + e_1 \end{aligned}$$

Let us solve for e_0. We can start by eliminating e_0 (while remembering that $e_0 = e_1$):

$$\begin{aligned} e_1 &= a(e_1 + e_2) + be_1 \\ e_2 &= \varepsilon + b(e_1 + e_1) + e_1 \\ &= \varepsilon + (b + \varepsilon)e_1 \end{aligned}$$

Let us now eliminate e_2:

$$\begin{aligned} e_1 &= a\bigl(e_1 + \varepsilon + (b + \varepsilon)e_1\bigr) + be_1 \\ &= (a + ab + b)e_1 + a \end{aligned}$$

Using Arden's lemma we get the unique solution

$$e_0 = e_1 = (a + ab + b)^* a = (a + b)^* a,$$

where the last step follows because ab is a member of the language generated by $(a + b)^*$. Thus the regular expression $e = (a + b)^* a$ satisfies $L(e) = L(A)$.

(b) It is easy to construct a DFA for the language generated by $(a+b)^* a$:

	a	b
$\rightarrow s_0$	s_1	s_0
$*s_1$	s_1	s_0

We can prove that this DFA is correct by converting it to a regular expression. It corresponds to the following system of equations, where e_0 corresponds to the start state s_0:

$$\begin{aligned} e_0 &= ae_1 + be_0 \\ e_1 &= \varepsilon + ae_1 + be_0 \end{aligned}$$

Let us solve for e_0. We can start by eliminating e_1 using Arden's lemma:

$$\begin{aligned} e_1 &= a^*(\varepsilon + be_0) \\ e_0 &= a\bigl(a^*(\varepsilon + be_0)\bigr) + be_0 \\ &= a^+(\varepsilon + be_0) + be_0 \\ &= a^+ + (a^+ + \varepsilon)be_0 \\ &= a^+ + a^* be_0 \end{aligned}$$

Using Arden's lemma again we get the (unique) solution

$$\begin{aligned} e_0 &= (a^* b)^* a^+ \\ &= (a^* b)^* a^* a \\ &= (a + b)^* a, \end{aligned}$$

where the final step uses the denesting rule.

Let us now minimise this DFA. First note that all of its states are accessible. Furthermore each state is distinguishable from every other: there are only two states, and only one of them is accepting. Thus the DFA is already minimal.

(c) The language $L(A)$ is equal to $L((a+b)^*a)$. Thus $L(A)$ consists of arbitrary (finite) strings of a's and/or b's, with the restriction that the strings have to end with an a.

4. No, they do not. We have

$$\begin{array}{ll} e_1 & = \\ ((\varepsilon + b)a)^*(ab(a+ba)^*)^* & = \\ (a+ba)^*(ab(a+ba)^*)^* & = \{\text{by the denesting rule}\} \\ (a+ba+ab)^*, & \end{array}$$

and

$$\begin{array}{ll} e_2 & = \\ (a+ba+ab)^*(b(a+ba+ab)^*)^* & = \{\text{by the denesting rule}\} \\ (a+b+ba+ab)^*. & \end{array}$$

Note that $b \in L(e_2) \setminus L(e_1)$.

5. (a) The grammar is $G = (\{\, \overline{X}, \overline{Y} \,\}, \{\, a, b \,\}, P, \overline{X})$, where the set of productions P contains exactly $\overline{X} \to a \mid a\overline{Y}$ and $\overline{Y} \to \overline{X}b$. See parts (b) and (c) for a proof showing that $L(G) = X$.

(b) The property $X \subseteq L(G)$ follows from $\forall w \in X.\ w \in L(G, \overline{X})$. Let us prove the latter statement, mutually with $\forall w \in Y.\ w \in L(G, \overline{Y})$, by induction on the structure of X and Y. We have three cases to consider:

- A case corresponding to $\dfrac{}{a \in X}$. In this case we should prove that $a \in L(G, \overline{X})$. We can construct the following derivation:

$$\dfrac{\dfrac{}{\overline{X} \to a \in P} \qquad \dfrac{\dfrac{}{\varepsilon \in L_{\mathrm{L}}(G, \varepsilon)}}{a \in L_{\mathrm{L}}(G, a)}}{a \in L(G, \overline{X})}$$

  (Antecedents of the form "a is a terminal" or "A is a nonterminal" are omitted from this and subsequent derivations.)

- A case corresponding to $\dfrac{w \in Y}{aw \in X}$. In this case we should prove that $aw \in L(G, \overline{X})$, given the inductive hypothesis that $w \in$

$L(G, \overline{Y})$. We can construct the following derivation:

$$\dfrac{\dfrac{}{\overline{X} \to a\overline{Y} \in P} \qquad \dfrac{\dfrac{w \in L(G, \overline{Y}) \quad \dfrac{}{\varepsilon \in L_{\mathrm{L}}(G, \varepsilon)}}{w \in L_{\mathrm{L}}(G, \overline{Y})}}{aw \in L_{\mathrm{L}}(G, a\overline{Y})}}{aw \in L(G, \overline{X})}$$

- A case corresponding to $\dfrac{w \in X}{wb \in Y}$. In this case we should prove that $wb \in L(G, \overline{Y})$, given the inductive hypothesis that $w \in L(G, \overline{X})$. We can construct the following derivation:

$$\dfrac{\dfrac{}{\overline{Y} \to \overline{X}b \in P} \qquad \dfrac{w \in L(G, \overline{X}) \quad \dfrac{\dfrac{}{\varepsilon \in L_{\mathrm{L}}(G, \varepsilon)}}{b \in L_{\mathrm{L}}(G, b)}}{wb \in L_{\mathrm{L}}(G, \overline{X}b)}}{wb \in L(G, \overline{Y})}$$

(c) The property $L(G) \subseteq X$ follows from

$$\forall w \in L(G, \overline{X}).\ w \in X. \tag{1}$$

Let us prove this, mutually with

$$\forall w \in L(G, \overline{Y}).\ w \in Y, \tag{2}$$

by complete induction on the lengths of the strings. We can begin by proving the first statement (**??**). The derivation of $w \in L(G, \overline{X})$ must end in the following way:

$$\dfrac{\overline{X} \to \alpha \in P \quad w \in L_{\mathrm{L}}(G, \alpha)}{w \in L(G, \overline{X})}$$

There are two possibilities for α:

- $\alpha = a$: In this case the derivation must end in the following way, and w must be equal to a:

$$\dfrac{\dfrac{}{\overline{X} \to a \in P} \quad \dfrac{\dfrac{}{\varepsilon \in L_{\mathrm{L}}(G, \varepsilon)}}{a \in L_{\mathrm{L}}(G, a)}}{a \in L(G, \overline{X})}$$

We can easily construct a derivation showing that $w = a \in X$:

$$\dfrac{}{a \in X}$$

- $\alpha = a\overline{Y}$: In this case the derivation must end in the following way, and w must be equal to au for some $u \in L(G, \overline{Y})$:

$$\dfrac{\dfrac{}{\overline{X} \to a\overline{Y} \in P} \qquad \dfrac{\dfrac{u \in L(G,\overline{Y}) \quad \dfrac{}{\varepsilon \in L_{\mathrm{L}}(G,\varepsilon)}}{u \in L_{\mathrm{L}}(G,\overline{Y})}}{au \in L_{\mathrm{L}}(G,a\overline{Y})}}{au \in L(G,\overline{X})}$$

  Note that $|u| < |w|$. One of the inductive hypotheses thus implies that $u \in Y$. We can now construct a derivation showing that $w = au \in X$:

$$\dfrac{u \in Y}{au \in X}$$

Let us now prove the second statement (**??**). The derivation of $w \in L(G, \overline{Y})$ must end in the following way, and w must be equal to ub for some $u \in L(G, \overline{X})$:

$$\dfrac{\dfrac{}{\overline{Y} \to \overline{X}b \in P} \qquad \dfrac{u \in L(G,\overline{X}) \quad \dfrac{\dfrac{}{\varepsilon \in L_{\mathrm{L}}(G,\varepsilon)}}{b \in L_{\mathrm{L}}(G,b)}}{ub \in L_{\mathrm{L}}(G,\overline{X}b)}}{ub \in L(G,\overline{Y})}$$

Note that $|u| < |w|$. One of the inductive hypotheses thus implies that $u \in X$. We can conclude by constructing a derivation showing that $w = ub \in Y$:

$$\dfrac{u \in X}{ub \in Y}$$

6. (a) Let us denote the language by L. It is equal to $\overline{M}\ \overline{N}$, where the complements are taken with respect to the language $\{\, a, b \,\}^*$,

$$\begin{aligned} M &= \left\{\, u \in \{\, a, b \,\}^* \;\middle|\; |u| < 3 \,\right\}, \text{ and} \\ N &= \left\{\, v \in \{\, a, b \,\}^* \;\middle|\; \exists v_1, v_2 \in \{\, a, b \,\}^* .\ v = v_1 a b v_2 \,\right\} \\ &= \left\{\, v_1 a b v_2 \;\middle|\; v_1, v_2 \in \{\, a, b \,\}^* \,\right\} \\ &= \{\, a, b \,\}^* \{\, ab \,\} \{\, a, b \,\}^* . \end{aligned}$$

The language M is regular because it is finite (and a finite language $\{\, w_1, ..., w_n \,\}$, where $n \in \mathbb{N}$, is regular because it is generated by the regular expression $w_1 + \cdots + w_n$). N is regular because it is generated by the regular expression $(a+b)^* ab (a+b)^*$. Finally the set of regular languages is closed under complementation (with respect to $\{\, a, b \,\}^*$) and concatenation, so $L = \overline{M}\ \overline{N}$ is regular. Every regular language is context-free, so L is also context-free.

(b) Let us denote the language by L. Consider the following function:

$$\begin{aligned}
&h \in \{\, a, b, c \,\} \to \{\, a, b \,\}^* \\
&h(a) = a \\
&h(b) = b \\
&h(c) = \varepsilon
\end{aligned}$$

The set of context-free languages is closed under string homomorphisms, so if L were context-free, then $h(L)$ would also be context-free. However,

$$\begin{aligned}
h(L) &= h\big(\big\{\, wcw \;\big|\; w \in \{\, a, b, c \,\}^*, \exists u, v \in \{\, a, b \,\}^*.\ w = ucv \,\big\}\big) \\
&= \big\{\, h(wcw) \;\big|\; w \in \{\, a, b, c \,\}^*, \exists u, v \in \{\, a, b \,\}^*.\ w = ucv \,\big\} \\
&= \big\{\, h(w)h(w) \;\big|\; w \in \{\, a, b, c \,\}^*, \exists u, v \in \{\, a, b \,\}^*.\ w = ucv \,\big\} \\
&= \big\{\, ww \;\big|\; w \in \{\, a, b \,\}^* \,\big\},
\end{aligned}$$

which is not context-free. Thus L is not context-free. Because regular languages are context-free L is also not regular.