Sample solutions for the examination of Finite automata and formal languages (DIT321/DIT322/TMV027/TMV028) from 2020-08-19

Nils Anders Danielsson

Note that in some cases I have not explained "step by step" why a certain algorithm produces a certain result, even though students who took the exam were asked to do this.

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

   (b) The CYK table:

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

   (c) The grammar is ambiguous, because there are two distinct leftmost derivations of aaa:

$$
\begin{aligned}
S &\Rightarrow_{\mathrm{lm}} AA \Rightarrow_{\mathrm{lm}} aA \Rightarrow_{\mathrm{lm}} aAA \Rightarrow_{\mathrm{lm}} aaA \Rightarrow_{\mathrm{lm}} aaa \\
S &\Rightarrow_{\mathrm{lm}} AA \Rightarrow_{\mathrm{lm}} AAA \Rightarrow_{\mathrm{lm}} aAA \Rightarrow_{\mathrm{lm}} aaA \Rightarrow_{\mathrm{lm}} aaa
\end{aligned}
$$

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

$$
\begin{aligned}
Q &= \{ \mathit{start}, \mathit{zeros}, \mathit{ones}, \mathit{blank}, \mathit{accept} \} \\
\Gamma &= \{ 0, 1, \sqcup \}
\end{aligned}
$$

$$\begin{aligned}
&\delta \in Q \times \Gamma \rightharpoonup Q \times \Gamma \times \{\, \mathsf{L}, \mathsf{R} \,\} \\
&\delta(\mathit{start}, 0) \;= (\mathit{ones}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{start}, 1) \;= (\mathit{zeros}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{zeros}, 0) = (\mathit{zeros}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{zeros}, 1) = (\mathit{blank}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{ones}, 1) \;= (\mathit{ones}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{ones}, 0) \;= (\mathit{blank}, \sqcup, \mathsf{R}) \\
&\delta(\mathit{blank}, \sqcup) = (\mathit{accept}, \sqcup, \mathsf{R})
\end{aligned}$$

The machine always moves to the right. It accepts if and only if it finds

- a 0, zero or more repetitions of 1, and a 0, followed by a blank, or
- a 1, zero or more repetitions of 0, and a 1, followed by a blank.

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

$$\begin{aligned}
e_0 &= a(e_0 + e_1) + be_0 \\
e_1 &= be_2 \\
e_2 &= \varepsilon + be_2
\end{aligned}$$

Let us solve for e_0. Using Arden's lemma we get the unique solution $e_2 = b^*$:

$$\begin{aligned}
e_0 &= a(e_0 + e_1) + be_0 \\
e_1 &= bb^* = b^+
\end{aligned}$$

Let us now eliminate e_1:

$$e_0 = a(e_0 + b^+) + be_0 = (a + b)e_0 + ab^+$$

Using Arden's lemma we get the unique solution $e_0 = (a + b)^*ab^+$. Thus the regular expression $e = (a + b)^*ab^+$ satisfies $L(e) = L(A)$.

(b) If the NFA A is converted to a DFA using the subset construction (with inaccessible states omitted), then we obtain the following DFA (possibly with different names for the states):

	a	b
$\rightarrow \{\, s_0 \,\}$	$\{\, s_0, s_1 \,\}$	$\{\, s_0 \,\}$
$\{\, s_0, s_1 \,\}$	$\{\, s_0, s_1 \,\}$	$\{\, s_0, s_2 \,\}$
$* \{\, s_0, s_2 \,\}$	$\{\, s_0, s_1 \,\}$	$\{\, s_0, s_2 \,\}$

Let us now minimise this DFA. Note first that all of its states are accessible. If the algorithm from the course is used to find equivalent states, then we see that every state of this DFA is distinguishable from every other state. Thus the DFA is already minimal.

(c) The language $L(A)$ is equal to $L((a + b)^*ab^+)$. Thus the language consists of all strings of the following form:

- First an arbitrary string consisting of a's and b's,
- then exactly one a,
- and finally one or more b's.

4. Let us start by converting the three ε-NFAs to equivalent DFAs by using the subset construction (with inaccessible states omitted):

(a)

	a	b
$\rightarrow s_0$	s_1	s_2
s_1	s_1	s_3
$*s_2$	s_4	s_5
$*s_3$	s_5	s_5
s_4	s_5	s_2
s_5	s_5	s_5

(b)

	a	b
$\rightarrow s_0$	s_1	s_2
$*s_1$	s_3	s_2
s_2	s_4	s_5
s_3	s_3	s_2
$*s_4$	s_5	s_2
s_5	s_5	s_5

(c)

	a	b
$\rightarrow s_0$	s_1	s_2
s_1	s_1	s_3
$*s_2$	s_4	s_5
$*s_3$	s_5	s_5
s_4	s_5	s_2
s_5	s_5	s_5

The first and last ε-NFAs were converted to equal DFAs, so they denote the same language. We can also see that the first and second ε-NFAs do not denote the same language, because the second one accepts the string a, which is not accepted by the first one.

5. (a) The grammar is $G = (\{\,\overline{X}, \overline{Y}\,\}, \{\,a, b\,\}, P, \overline{X})$, where the set of productions P contains exactly $\overline{X} \rightarrow a \mid \overline{X} b \overline{Y}$ and $\overline{Y} \rightarrow a \overline{X}\,\overline{X}$. 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 $\dfrac{u \in X \quad v \in Y}{ubv \in X}$. In this case we should prove that $ubv \in L(G, \overline{X})$, given the inductive hypotheses that $u \in L(G, \overline{X})$ and $v \in L(G, \overline{Y})$. We can construct the following derivation:

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

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

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

(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 (1). 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:

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

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

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

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

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

  Note that $|u| < |w|$ and $|v| < |w|$. The inductive hypotheses thus imply that $u \in X$ and $v \in Y$. We can now construct a derivation showing that $w = ubv \in X$:

$$\dfrac{u \in X \quad v \in Y}{ubv \in X}$$

Let us now prove the second statement (2). The derivation of $w \in L(G, \overline{Y})$ must end in the following way, and w must be equal to auv for some $u, v \in L(G, \overline{X})$:

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

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

$$\dfrac{u, v \in X}{auv \in Y}$$

6. (a) Let us denote the language by L. It is equal to MN, where

$$M = \left\{ u \mid u \in \{ a, b \}^* \right\} = \{ a, b \}^*, \text{ and}$$
$$N = \left\{ vv \mid v \in \{ a, b \}^*, |v| \leq 3 \right\}.$$

The language M is regular because it is generated by the regular expression $(a + b)^*$. N 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 $w_1 + \cdots + w_n$). Finally the set of regular languages is closed under concatenation, so $L = MN$ is regular. Every regular language is context-free, so L is also context-free.

(b) Let us denote the language by M. It is equal to $\{ c \} L$, where

$$L = \left\{ ww^{\mathrm{R}} \mid w \in \{ a, b, c \}^* \right\}.$$

Both $\{ c \}$ and L are context-free. Thus, because the set of context-free languages is closed under concatenation, we get that $M = \{ c \} L$ is context-free.

We also have that L is not regular. Thus it follows from the following lemma that M is not regular.

Lemma. *If Σ is an alphabet with $a \in \Sigma$, $L \subseteq \Sigma^*$, and $\{ av \mid v \in L \}$ is regular, then L is regular.*

Proof. Take a DFA $D = (Q, \Sigma, \delta, q_0, F)$ for $\{ av \mid v \in L \}$. We can construct a DFA $D' = (Q, \Sigma, \delta, \delta(q_0, a), F)$, for which we have

$$\begin{aligned} L(D') &= \left\{ w \in \Sigma^* \;\middle|\; \hat{\delta}(\delta(q_0, a), w) \in F \right\} \\ &= \left\{ w \in \Sigma^* \;\middle|\; \hat{\delta}(q_0, aw) \in F \right\} \\ &= \{ w \in \Sigma^* \mid aw \in \{ av \mid v \in L \} \} \\ &= L. \end{aligned}$$

Thus L is regular, because it is the language of the DFA D'. $\square$