ToC Spring 2022

Practice Problems -1

1. Textbook, Page 83, Exercise 1.4.

   Draw **DFA** for:

   (a) i. $\{w|\ w$ has at least three a's$\}$

   ii. $\{w|\ w$ has at least two b's$\}$

   iii. $\{w|\ w$ has at least three a's and at least two b's$\}$

   (b) i. $\{w|\ w$ has exactly two a's$\}$

   ii. $\{w|\ w$ has at least two b's$\}$

   iii. $\{w|\ w$ has exactly two a's and at least two b's$\}$

   (c) i. $\{w|\ w$ has an even number of a's$\}$

   ii. $\{w|\ w$ has one or two b's$\}$

   iii. $\{w|\ w$ has an even number of a's and one or two b's$\}$

   (d) i. $\{w|\ w$ has an even number of a's$\}$

   ii. $\{w|$ each a in w is followed by at least one b$\}$

   iii. $\{w|\ w$ has an even number of a's and each a is followed by at least one b$\}$

   (e) i. $\{w|\ w$ starts with an a$\}$

   ii. $\{w|\ w$ has at most one b$\}$

   iii. $\{w|\ w$ starts with an a and has at most one b$\}$

(f) i. $\{w \mid w$ has an odd number of a's$\}$

ii. $\{w \mid w$ ends with a b$\}$

iii. $\{w \mid w$ has an odd number of a's and ends with a b$\}$

(g) i. $\{w \mid w$ has even length$\}$

ii. $\{w \mid w$ has an odd number of a's$\}$

iii. $\{w \mid w$ has even length and an odd number of a's$\}$

2. Textbook, Page 84, Exercise 1.7. Draw **NFA** for:

(a) The language $\{w \mid w$ ends with 00$\}$ with three states
(b) The language $\{0\}$ with two states
(c) The language $0^*1^*0^+$ with three states
(d) The language $1^*(001^+)^*$ with three states
(e) The language $\{\varepsilon\}$ with one state
(f) The language 0^* with one state

3. Write the **RE** for:
(a) $\{w \mid w$ begins with a 1 and ends with a 0$\}$

(b) $\{w \mid w$ contains at least three 1s$\}$

(c) $\{w \mid w$ contains the substring 0101, i.e., $w = x0101y$ for some x and $y\}$

(d) $\{w \mid w$ has length at least 3 and its third symbol is a 0$\}$

(e) $\{w \mid w$ starts with 0 and has odd length, or starts with 1 and has even length$\}$

(f) $\{w \mid w$ doesn't contain the substring 110$\}$

Practice Assignment 2

1. Give a context-free grammar (CFG) for each of the following languages over the alphabet $\Sigma = \{a,b\}$:

   (a) All strings in the language $L : \{a^n b^m a^{2n} | n,m \geq 0\}$

   (b) All nonempty strings that start and end with the same symbol.

   (c) All strings with more a's than b's.

2. Consider the grammar below that generates roman numerals, with terminals $\{\mathbf{c},\mathbf{l},\mathbf{x},\mathbf{v},\mathbf{i}\}$. c = 100,l = 50,x = 10,v = 5,i = 1. Notice that we use lowercase characters here to represent the numerals, to distinguish them from the non-terminals.

$$
\begin{aligned}
S &\rightarrow \mathbf{x}TU \mid \mathbf{l}X \mid X \\
T &\rightarrow \mathbf{c} \mid \mathbf{l} \\
X &\rightarrow \mathbf{x}X \mid U \\
U &\rightarrow \mathbf{i}Y \mid \mathbf{v}I \mid I \\
Y &\rightarrow \mathbf{x} \mid \mathbf{v} \\
I &\rightarrow \mathbf{i}I \mid \epsilon
\end{aligned}
$$

   (a) Draw a parse tree for 47: "xlvii".

   (b) Is this grammar ambiguous?

3. Let L be the language $\{w \in \{a,b\}^* | w$ contains exactly one more b than $a\}$.

   (a) Give a context-free grammar that generates L.

   (b) Give a leftmost derivation and a parse tree in your grammar for the string $abbabab$.

   (c) Give an unambiguous grammar for the language L of the previous problem.

   (This is a difficult problem, but give it a try. As a hint, you can use three variables other than the start symbol. One variable generates strings with the same number of a's as b's, the second variable generates strings with the same number of a's as b's that have the additional property that every prefix has at least as many a's as b's, and the third variable generates all strings with the same number of a's as b's that have the additional property that every prefix has at least as many b's as a's.)

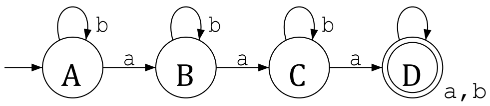
4. Prove $L_4 = \{a^i b^{2i} c^i \mid i > 0\}$ is not a Context-free Language

5. Prove $L_5 = \{s2s \mid s \epsilon \{0,1\}^*\}$ is not a Context-free Language

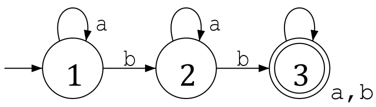
Practice 1 – Solutions

1. Textbook, Page 83, Exercise 1.4. (a)

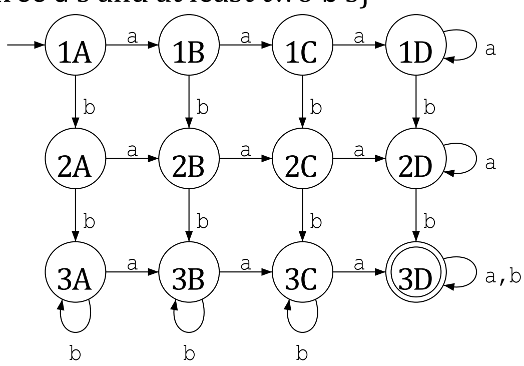
i. $\{w \mid w$ has at least three a's$\}$



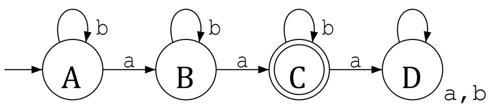
ii. $\{w \mid w$ has at least two b's$\}$



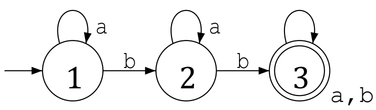
iii. $\{w \mid w$ has at least three a's and at least two b's$\}$



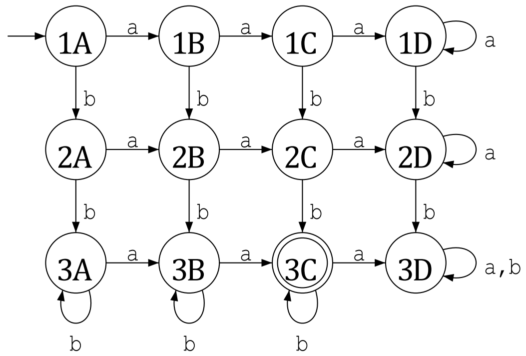
(b) i. $\{w \mid w$ has exactly two a's$\}$



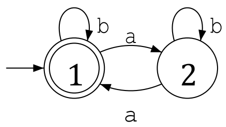
ii. $\{w \mid w$ has at least two b's$\}$



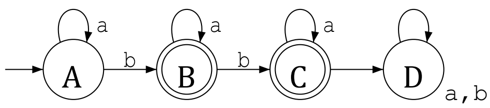
iii. $\{w \mid w$ has exactly two a's and at least two b's$\}$



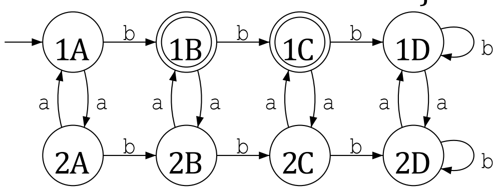
(c) i. $\{w \mid w$ has an even number of a's$\}$



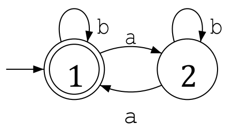
ii. $\{w \mid w$ has one or two b's$\}$



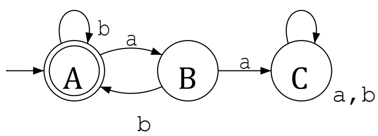
iii. $\{w \mid w$ has an even number of a's and one or two b's$\}$



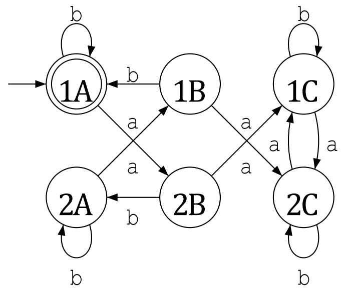
(d) i. $\{w \mid w$ has an even number of a's$\}$



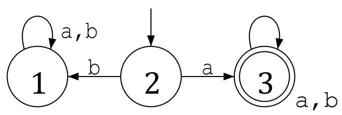
ii. $\{w \mid$ each a in w is followed by at least one b$\}$



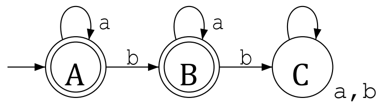
iii. $\{w \mid w$ has an even number of a's and each a is followed by at least one b$\}$



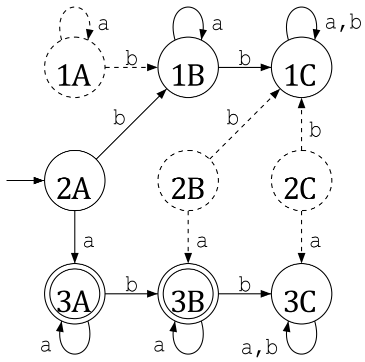
(e) i. $\{w \mid w$ starts with an a$\}$



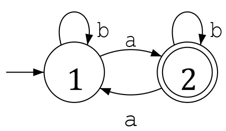
ii. $\{w \mid w$ has at most one b$\}$



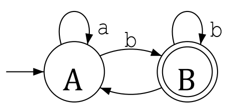
iii. $\{w \mid w$ starts with an a and has at most one b$\}$



(f) i. $\{w \mid w$ has an odd number of a's$\}$

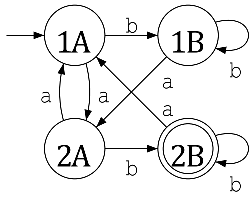


ii. $\{w \mid w$ ends with a b$\}$

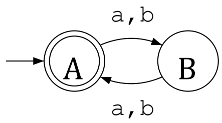


a

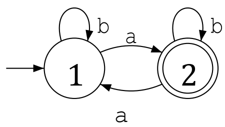
iii. {w| w has an odd number of a’s and ends with a b}



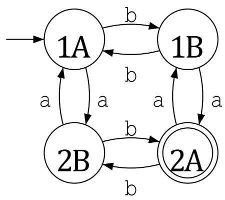
(g) i. {w| w has even length}



ii. {w| w has an odd number of a’s}

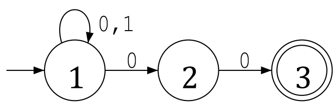


iii. {w| w has even length and an odd number of a’s}

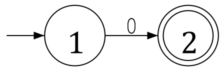


2. Textbook, Page 84, Exercise 1.7.

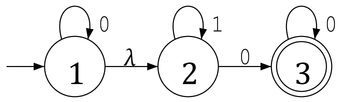
(a) The language {w| w ends with 00} with three states



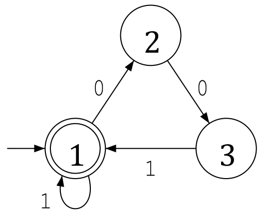
(b) The language {0} with two states



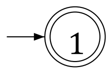
(c) The language $0^*1^*0^+$ with three states



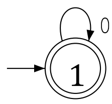
(d) The language $1^*(001^+)^*$ with three states



(e) The language $\{\varepsilon\}$ with one state



(f) The language 0^* with one state



3. Textbook, Page 86, Exercise 1.18. Give regular expressions generating the languages ofExercise 1.6.

   (a) $\{w|\ w$ begins with a 1 and ends with a 0$\}$

   $1\Sigma^*0$

   (b) $\{w|\ w$ contains at least three 1s$\}$

   $\Sigma_*1\Sigma_*1\Sigma_*1\Sigma_*$

   (c) $\{w|\ w$ contains the substring 0101, i.e., $w = x0101y$ for some x and $y\}$

$\Sigma^*0101\Sigma^*$

(d) $\{w|\ w$ has length at least 3 and its third symbol is a 0$\}$

$\Sigma\Sigma0\Sigma^*$

(e) $\{w|\ w$ starts with 0 and has odd length, or starts with 1 and has even length$\}$

$(0 \cup 1\Sigma)(\Sigma\Sigma)^*$

(f) $\{w|\ w$ doesn't contain the substring 110$\}$

$0_*(10^+)_*1_*$

Practice 2

Solutions

1. Give a context-free grammar (CFG) for each of the following languages over the alphabet $\Sigma = \{a,b\}$:

(a) All strings in the language $L : \{a^n b^m a^{2n} | n,m \geq 0\}$

$$\begin{aligned} S &\rightarrow aSaa \mid B \\ B &\rightarrow bB \mid \epsilon \end{aligned}$$

(b) All nonempty strings that start and end with the same symbol.

$$\begin{aligned} S &\rightarrow aXa \mid bXb \mid a \mid b \\ X &\rightarrow aX \mid bX \mid \epsilon \end{aligned}$$

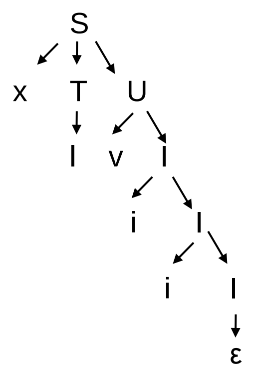
(c) All strings with more a's than b's.

$$\begin{aligned} S &\rightarrow Aa \mid MS \mid SMA \\ A &\rightarrow Aa \mid \epsilon \\ M &\rightarrow \epsilon \mid MM \mid bMa \mid aMb \end{aligned}$$

2. Consider the grammar below that generates roman numerals, with terminals $\{\mathbf{c},\mathbf{l},\mathbf{x},\mathbf{v},\mathbf{i}\}$. $c = 100, l = 50, x = 10, v = 5, i = 1$. Notice that we use lowercase characters here to represent the numerals, to distinguish them from the non-terminals.

$$\begin{aligned} S &\rightarrow \mathbf{x}TU \mid \mathbf{l}X \mid X \\ T &\rightarrow \mathbf{c} \mid \mathbf{l} \\ X &\rightarrow \mathbf{x}X \mid U \\ U &\rightarrow \mathbf{i}Y \mid \mathbf{v}I \mid I \\ Y &\rightarrow \mathbf{x} \mid \mathbf{v} \\ I &\rightarrow \mathbf{i}I \mid \epsilon \end{aligned}$$

(a) Draw a parse tree for 47: "xlvii".



(b) Is this grammar ambiguous?

No

3. Let L be the language $\{w \in \{a,b\}^* | w$ contains exactly one more b than $a\}$.

(a) Give a context-free grammar that generates L.

Solution:

$$\begin{aligned} S &\rightarrow \varepsilon|aB|bA \\ A &\rightarrow aS|bAA \\ B &\rightarrow bS|aBB \end{aligned}$$

we know that the variable B generates the strings with exactly one more b than a, so if we declare B to be the start symbol in the above grammar, we have one solution to the problem. To make this problem more interesting, I will give a different solution:

$$\begin{aligned} S &\rightarrow TbT \\ T &\rightarrow aTb|bTa|TT|\varepsilon \end{aligned}$$

You were not asked to explain how your grammar works, but here is an explanation for the above grammar. We know that the variable T generates the strings with the same number of a's and b's. Now suppose we have a string w with exactly one more b than a. Then, we have to show that w matches the rule $S \rightarrow TbT$. We think of a counter running along w where b counts as +1 and a counts as −1. The count at the end of w is +1, so we divide w up into ucv where c is the symbol read when the count first reaches +1. The count at the end of u is either 0 or 2, but if the count is 2 at the end of u, then the count must have been 1 somewhere in the middle of u, contradicting how we picked c, so the count is 0 at the end of u and c must be a b. Since u brings the counter from 0 to 0, u must have the same number of a's as b's, and since v brings the counter from 1 to 1, v also has the same number of a's as b's. Thus, $w = ubv$ matches the rule $S \rightarrow TbT$.

(b) Give a leftmost derivation and a parse tree in your grammar for thestring *abbabab*.

Solution: A leftmost derivation in the second grammar is: $S \Rightarrow TbT$

$$\Rightarrow aTbbT \Rightarrow abbT \Rightarrow abbaTb \Rightarrow abbabTab \Rightarrow abbabab$$

(c). Give an unambiguous grammar for the language L of the previous problem.

(This is a difficult problem, but give it a try. As a hint, you can use three variables other than the start symbol. One variable generates strings with the same number of a's as b's, the second variable generates strings with the same number of a's as b's that have the additional property that every prefix has at least as many a's as b's, and the third variable generates all strings with the same number of a's as b's that have the additional property that every prefix has at least as many b's as a's.)

Solution:

$$\begin{aligned} S &\rightarrow WbT \\ T &\rightarrow \varepsilon | aWbT | bVaT \\ W &\rightarrow \varepsilon | aWbW \\ V &\rightarrow \varepsilon | bVaV \end{aligned}$$

2. Prove L4 = $\{a^i b^{2i} c^i \mid i > 0\}$ is not a Context-free Language

Let us assume the language $\mathcal{L}_4$ is context-free. Then the Pumping lemma for context-free languages applies for $\mathcal{L}_4$.

Let n be the constant stated by the Pumping lemma.

Let $w = a^n b^{2n} c^n$. Clearly $w \in \mathcal{L}_4$ and $|w| \geqslant n$.

By the lemma we know that $w = xuyvz$ with $|uyv| \leqslant n$. We know also that $uv \neq \epsilon$, and hence we know that either u or v could be empty but not both.

We have five different possibilities for the location of uyv within w (recall we do not know where exactly uyv is located in the word):

1. uyv consists only of a's: then for any $k > 1$ the word xu^kyv^kz will have more a's than c's (because $uv \neq \epsilon$), and hence it will not belong to $\mathcal{L}_4$;
2. uyv consists only of b's: then for any $k > 1$ the word xu^kyv^kz will have more b's than a's and c's together (again because $uv \neq \epsilon$), and hence it will not belong to $\mathcal{L}_4$;
3. uyv consists only of c's: then for any $k > 1$ the word xu^kyv^kz will have more c's than a's (because $uv \neq \epsilon$), and hence it will not belong to $\mathcal{L}_4$;

4. uyv consists of a's and b's: recall that we know that $uv \neq \epsilon$ but we do not really know if both u and v are non-empty. So, for any $k > 1$, when pumping u and v we could actually only be pumping a's (if u is not empty and contains only a's, and v is empty), only pumping b's (if u is empty, and v is not empty and contains only b's), or we could be pumping both a's and b's (if neither u nor v are empty, and/or when at least one of them contains both a's and b's). We do know however that we are pumping at least one a or one b, and that we are not pumping any c's. So we know that xu^kyv^kz will not belong to $\mathcal{L}_4$ because we would have more a's than c's (whenever we pump a's), or more b's than a's and c's together (whenever we pump b'a but no a's). Observe that in the case u and/or v contain occurrences of both a's and b's, xu^kyv^kz will not even have the symbols in the right order when $k > 1$;

5. uyv consists of b's and c's: this is similar to the case before, only that for any $k > 1$ we will be pumping at least one b or one c, but we will not pump any a's. So we know that xu^kyv^kz will not belong to $\mathcal{L}_4$ because we would have more c's than a's (whenever we pump c's), or more b's than a's and c's together (whenever we pump b's but no c's). Even here, the order of the symbols in xu^kyv^kz might not be the right one if either u or v contain occurrences of both b's and c's when $k > 1$.

So, for each of these five possibilities there exists at least one k for which xu^kyv^kz does not belong to $\mathcal{L}_4$. Therefore $\mathcal{L}_4$ cannot be context-free.

Observe that in this example, uyv cannot consists of a's, b's AND c's. This is due to the fact that $|uyv| \leqslant n$ and also that w has $2n$ b's between the a's and the c's. Hence no part of the word containing at most n symbols could consists of all three letters.

3. Prove L5 = {s2s | sϵ{0,1}*} is not a Context-free Language

Let us assume the language $\mathcal{L}_5$ is context-free. Then the Pumping lemma for context-free languages applies for $\mathcal{L}_5$.

Let n be the constant given by the Pumping lemma.

Let $w = 0^n1^n20^n1^n$. Clearly $w \in \mathcal{L}_5$ and $|w| \geqslant n$.

By the lemma we know that $w = xuyvz$ with $|uyv| \leqslant n$. We know also that $uv \neq \epsilon$, and hence we know that either u or v could be empty but not both.

We have three different possibilities for the location of uyv within w.

If uyv takes place completely before the 2, then for $k = 0$ and since $uv \neq \epsilon$, whatever part of the word is removed before the 2 will not be removed after the 2. Hence xu^kyv^kz will not have the form $s2s$ and therefore it will not belong to $\mathcal{L}_5$.

We have a similar reasoning if uyv takes place completely after the 2.

If 2 is part of uyv then uyv should be of the form 1^i20^j for i, j such that $i + j + 1 \leqslant n$. Even here, for $k = 0$ and since $uv \neq \epsilon$, the resulting word is not of the form $s2s$ either because we remove the 2 (if 2 is part of uv), or because we remove 1's (right before the 2) and/or we remove 0's (right after the 2) but we do not remove 1's at the end of the word nor 0's at the beginning of the word. Hence xu^kyv^kz will not have the form $s2s$ and therefore it will not belong to $\mathcal{L}_5$.

So, for each of these three possibilities xu^0yv^0z does not belong to $\mathcal{L}_5$. Therefore $\mathcal{L}_5$ cannot be context-free.