TOC PROBLEM SET-8

SIDDHANT CHAUDHARY
BMC201953

1. State if the following languages are context free. Justify your answer. (Assume $\Sigma = \{a, b\}$).
(a) $L = \{a^n\#a^{2n}\#a^{3n}\}$. This is an easy application of the pumping lemma.
(b) $L = \{w\#x \mid w \text{ is a substring of } x\}$. This is context free. It is easy to make a PDA for this.

2. Given a CFL $\mathcal{L}$ and a regular language $\mathcal{R}$, let

$$\mathcal{L}_{\mathcal{R}} = \{w \mid w \in \mathcal{L}, \exists v \in \mathcal{R} \text{ such that } v \text{ is a subword of } w\}$$

Prove that $\mathcal{L}_{\mathcal{R}}$ is context free.

Solution. Here, we can use the fact that the intersection of a CFL with a regular language is a CFL. Let $\mathcal{R}'$ be the language of all words containing a subword in $\mathcal{R}$. We will show that $\mathcal{R}'$ is a regular language as well. Since $\mathcal{R}$ is regular, it has a regular expression, say R. Then, the expression

$$\Sigma^* R \Sigma^*$$

is clearly a regular expression for $\mathcal{R}'$, and hence $\mathcal{R}'$ is a regular language. Finally, it is easy to see that

$$\mathcal{L}_{\mathcal{R}} = \mathcal{L} \cap \mathcal{R}'$$

and hence we conclude that $\mathcal{L}_{\mathcal{R}}$ is a regular langauge, and this completes the proof.

3. Given a context free grammar G, describe a procedure to check if $L(G)$ is finite.
Bonus: What can you say about the complexity of your procedure?

Solution. To be completed

4. Convert the following CFG into a CFG in Chomsky Normal Form (which accepts all words in the given grammar except ϵ), using the procedure described in the lectures.

$$S \to TU$$
$$T \to aTb \mid \epsilon$$
$$U \to bTa \mid \epsilon$$

Date: October 2020.

Solution. First, we can introduce non-terminals for the symbols a, b, giving us the following grammar.

$$
\begin{aligned}
S &\to TU \\
T &\to ATB \mid \epsilon \\
U &\to BTA \mid \epsilon \\
A &\to a \\
B &\to b
\end{aligned}
$$

Next, we introduce new non-terminals as follows.

$$
\begin{aligned}
S &\to TU \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \\
A &\to a \\
B &\to b
\end{aligned}
$$

Next, to remove $T \to \epsilon$, new productions will be introduced as follows.

$$
\begin{aligned}
S &\to TU \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \\
A &\to a \\
B &\to b \\
S &\to U \\
X_1 &\to B \\
X_2 &\to A
\end{aligned}
$$

Next, to remove $U \to \epsilon$, a new production will be added as follows.

$$
\begin{aligned}
S &\to TU \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \\
A &\to a \\
B &\to b \\
S &\to U \\
X_1 &\to B \\
X_2 &\to A \\
S &\to T
\end{aligned}
$$

Now, we move to unit productions. To handle $S \to U$, the following productions to S will be added.

$$
\begin{aligned}
S &\to TU \mid BX_2 \mid \epsilon \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \\
A &\to a \\
B &\to b \\
S &\to U \\
X_1 &\to B \\
X_2 &\to A \\
S &\to T
\end{aligned}
$$

To handle the productions $X_1 \to B, X_2 \to A$, the productions $X_1 \to b, X_2 \to a$ will be added to give:

$$
\begin{aligned}
S &\to TU \mid BX_2 \mid \epsilon \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \mid b \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \mid a \\
A &\to a \\
B &\to b \\
S &\to U \\
X_1 &\to B \\
X_2 &\to A \\
S &\to T
\end{aligned}
$$

Finally, to handle $S \to T$, the following will be added.

$$
\begin{aligned}
S &\to TU \mid BX_2 \mid AX_1 \mid \epsilon \\
T &\to AX_1 \mid \epsilon \\
X_1 &\to TB \mid b \\
U &\to BX_2 \mid \epsilon \\
X_2 &\to TA \mid a \\
A &\to a \\
B &\to b \\
S &\to U \\
X_1 &\to B \\
X_2 &\to A \\
S &\to T
\end{aligned}
$$

Finally, we remove all ϵ and unit productions to get the Chomsky Normal Form, which is below.

$$
\begin{aligned}
S &\to TU \mid BX_2 \mid AX_1 \\
T &\to AX_1 \\
X_1 &\to TB \mid b \\
U &\to BX_2 \\
X_2 &\to TA \mid a \\
A &\to a \\
B &\to b
\end{aligned}
$$

In the following two problems, prove/disprove whether context free languages are closed under the given operations.

5. DROP$(\mathcal{L}) = \{xz \mid xyz \in \mathcal{L}, y \in \Sigma, x, z \in \Sigma^*\}$.

Solution. It is true that CFLs are closed under the DROP operation, which I will now prove. Let $G = (N, \Sigma, P, X_1)$ be a context free grammar. Suppose $X_1, ..., X_k$ are the non-terminals of G, where the starting non-terminal is X_1. We will make a new grammar G' that will accept the language DROP$(L(G))$ as follows. For every $1 \leq i \leq k$, add a new non-terminal X_i^{copy} to the new grammar G'. The new non terminal X_i^{copy} will act like the non-terminal X_i, just that the copy will have the *potential* to drop a letter, and we formalize this below.

Let the starting non-terminal of G' be X_1^{copy} (note that X_1 was the starting non-terminal for G). Now we describe the productions in G'. Let G' have all the productions of G, and we add new productions as follows. Let

$$X_i \to \alpha$$

be a production in G where $\alpha \in (N \cup \Sigma)^*$. Let $s \in \Sigma$ be a letter in α, i.e we can write $\alpha = \alpha_s s \beta_s$, where $\alpha_s, \beta_s \in (N \cup \Sigma)^*$. Then, for *every* such letter s, add the production

$$X_i^{\text{copy}} \to \alpha_s \beta_s$$

in the new grammar G' (these productions help in dropping exactly one letter). Next, suppose

$$X_i \to \alpha$$

is a production in G where $\alpha \in (N \cup \Sigma)^*$. Let $X_j \in N$ be a non-terminal in α, i.e we can write $\alpha = \alpha_{X_j} X_j \beta_{X_j}$, where $\alpha_{X_j}, \beta_{X_j} \in (N \cup \Sigma)^*$. Then, for *every* such X_j, add the production

$$X_i^{\text{copy}} \to \alpha_{X_j} X_j^{\text{copy}} \beta_{X_j}$$

to the new grammar G' (these productions ensure that if X_i^{copy} does not drop a letter, then it has a production containing a non-terminal that can potentially drop). Observe that these productions ensure that *only* one non-terminal drops, and that *atleast* one non-terminal trops. So this shows that $L(G') =$ DROP$(L(G))$, showing that CFLs are closed under this operation.

6. ROC$(\mathcal{L}) = \{yx \mid xy \in \mathcal{L}\}$.

Solution. To be completed