How to write proofs - II

Problem (Ullman and Hopcroft, Page – 104, Exercise – 4.8)**.** *Let G be the grammar*

$$\begin{aligned} S &\rightarrow aB \mid bA \\ A &\rightarrow a \mid aS \mid bAA \\ B &\rightarrow b \mid bS \mid aBB \end{aligned}$$

Prove that

$$L(G) = \left\{ \; x \in \{a,b\}^* \; \mid x \text{ contains equal number of } a\text{'s and } b\text{'s} \; \right\}.$$

Proof. To prove the statement above, we would prove the following lemma first.

Lemma 1. *If x is a non-empty string over $\{a,b\}$, then the following statements hold –*

(a) If x contains equal number of a's and b's, then $S \stackrel{*}{\Rightarrow} x$.

(b) If x contains one more a than b, then $A \stackrel{*}{\Rightarrow} x$.

(c) If x contains one more b than a, then $B \stackrel{*}{\Rightarrow} x$.

Proof of lemma 1. We will prove the lemma by induction on the length of x. Possible non-empty strings over $\{a,b\}$ of length less than or equal to 2 satisfying one of the above criteria are

$$\{a, b, ab, ba\}.$$

If $x = a$, then

$$A \Rightarrow a.$$

If $x = b$, then

$$B \Rightarrow b.$$

If $x = ab$, then

$$S \Rightarrow aB \Rightarrow ab.$$

If $x = ba$, then

$$S \Rightarrow bA \Rightarrow ba.$$

So, the statements hold when $1 \leq |x| \leq 2$.

Let us assume that the statements hold for all x such that $|x| \leq k$ for some integer $k \geq 1$, and x satisfies one of the above three criteria. Then, consider a string y of length $k+1$. Depending on the composition of y, we have 3 cases –

(I) y has equal number of a's and b's. Let the first character of y be a. So,

$$y = ay',$$

where

(i) $|y'| = k$, and

(ii) number of b's in y' is one more than the number of a's.

By induction hypothesis,

$$B \stackrel{*}{\Rightarrow} y'.$$

So,

$$S \Rightarrow aB \stackrel{*}{\Rightarrow} ay' = y.$$

On the other hand, if the first character of y is b, then y can be expressed as

$$y = by'$$

where

(i) $|y'| = k$, and

(ii) number of a's in y' is one more than the number of b's.

So, by induction hypothesis

$$A \stackrel{*}{\Rightarrow} y'.$$

So,

$$S \Rightarrow bA \stackrel{*}{\Rightarrow} by' = y.$$

(II) y has one more a than b. Let the first character of y be a. So, y can be expressed as

$$y = ay',$$

where

(i) $|y'| = k$, and

(ii) number of a's and b's in y' are equal.

So, by induction hypothesis

$$S \stackrel{*}{\Rightarrow} y'.$$

Hence,

$$A \Rightarrow aS \stackrel{*}{\Rightarrow} ay' = y.$$

On the other hand, let the first character of y be b. So, y can be expressed as

$$y = by',$$

where

(i) $|y'| = k$, and

(ii) number of a's in y' is two more than the number of b's.

Now, we can observe that the difference in the count of a's and b's in y' at the start of the string is 0, and at the end is 2. So, there must be a prefix of y', say u, such that the number of a's in u is one more than the number of b's. So, y' can be exressed as

$$y' = uv,$$

where $|u| < |y'|, |v| < |y'|$ and the number of a's in u and in v is one more than the number of b's. We also note that $|u| < k, |v| < k$. So, by induction hypothesis

$$\begin{aligned} A &\stackrel{*}{\Rightarrow} u, \text{ and} \\ A &\stackrel{*}{\Rightarrow} v. \end{aligned}$$

So,

$$A \Rightarrow bAA \stackrel{*}{\Rightarrow} buA \stackrel{*}{\Rightarrow} buv = y.$$

(III) y has one more b than a. This case can be handled in an analogous manner to the case above with one difference. In this case, we can show that

$$B \stackrel{*}{\Rightarrow} y.$$

Finally, with the 3 cases established for y of length $k+1$ satisfying the criteria of the lemma, we have proved the statements of the lemma for all non-empty strings.

□

We now prove a result that shows the other direction of lemma 1.

Lemma 2. *The following statements hold for non-empty strings over* $\{a, b\}$.

(a) If $S \stackrel{*}{\Rightarrow} x$*, then* x *contains equal number of* a*'s and* b*'s.*

(b) If $A \stackrel{*}{\Rightarrow} x$*, then* x *contains one more* a *than* b*.*

(c) If $B \stackrel{*}{\Rightarrow} x$*, then* x *contains one more* b *than* a*.*

Proof of lemma 2. We will prove the lemma by induction on the number of steps is takes to derive x.

When the number of derivation steps is ≤ 2, S, A and B derive the following strings –

$$S \Rightarrow aB \Rightarrow ab$$

$$S \Rightarrow bA \Rightarrow ba$$

$$A \Rightarrow a$$

$$B \Rightarrow b$$

So, the statements hold when the number of derivation steps are ≤ 2.

Let us assume that the statements hold for all strings that can be derived from S, A and B in $\leq k$ steps. Then, consider the strings that are derived in $k+1$ steps.

(I) $S \stackrel{*}{\Rightarrow} x$ in $k+1$ steps. So,

$$S \stackrel{k+1}{\Longrightarrow} x$$

In the above derivation, consider the first derivation step. Based on the given grammar, there are two possibilites –

(a) The first derivation step uses the production $S \to aB$. In that case

$$S \Rightarrow aB \stackrel{k}{\Longrightarrow} x$$

So, the first symbol of x is a. Expressing x as

$$x = ax',$$

we can rewrite the derivation as

$$S \Rightarrow aB \stackrel{k}{\Longrightarrow} ax'.$$

Thus,

$$B \stackrel{k}{\Longrightarrow} x'.$$

By induction hypothesis, x' has one more b than a. So, x, which has been expressed as ax' has equal number of a's and b's.

(b) The first derivation step uses the production $S \to bA$. In that case

$$S \Rightarrow bA \stackrel{k}{\Longrightarrow} x$$

So, the first symbol of x is b. Expressing x as

$$x = bx',$$

we can rewrite the derivation as

$$S \Rightarrow bA \stackrel{k}{\Longrightarrow} bx'.$$

Thus,

$$A \stackrel{k}{\Longrightarrow} x'.$$

By induction hypothesis, x' has one more a than b. So, x, which has been expressed as bx' has equal number of a's and b's.

(II) $A \stackrel{*}{\Rightarrow} x$ in $k+1$ steps. So,

$$A \stackrel{k+1}{\Longrightarrow} x$$

In the above derivation, consider the first derivation step. Based on the given grammar, there are two possibilites –

(a) The first derivation step uses the production $A \to aS$. In that case

$$A \Rightarrow aS \overset{k}{\Longrightarrow} x$$

So, the first symbol of x is a. Expressing x as

$$x = ax',$$

we can rewrite the derivation as

$$A \Rightarrow aS \overset{k}{\Longrightarrow} ax'.$$

Thus,

$$S \overset{k}{\Longrightarrow} x'.$$

By induction hypothesis, x' has equal number of a's and b's. So, x, which has been expressed as ax' has one more a than b.

(b) The first derivation step uses the production $A \to bAA$. In that case

$$A \Rightarrow bAA \overset{k}{\Longrightarrow} x$$

So, the first symbol of x is b. Expressing x as

$$x = bx',$$

we can rewrite the derivation as

$$A \Rightarrow bAA \overset{k}{\Longrightarrow} bx'.$$

Let the first of the A's in bAA derive the string u and the second of the A derive the string v. So, the derivation can be rewritten as

$$A \Rightarrow bAA \overset{<k}{\Longrightarrow} buA \overset{<k}{\Longrightarrow} buv = bx'.$$

So, we have the following observations –

i. $x = bx' = buv$,

ii. $A \overset{<k}{\Longrightarrow} u$, and

iii. $A \overset{<k}{\Longrightarrow} v$.

Thus, by induction hypothesis, both u and v have the property that the number of a's in them is one more than the number of b's. So, in $x = buv$, the number of a's is one more than the number of b's.

(III) $B \overset{*}{\Rightarrow} x$ in $k+1$ steps. This case can be handled in a manner analogous to the earlier case.

Finally, with the three cases established for derivation steps of length $k+1$, we have proved the statements of the lemma for all non-emplty strings.

□

Combining lemmas 1 and 2, we have the following theorem.

Theorem 1. *The following statements hold for non-empty strings over* $\{a, b\}$.

(a) $S \overset{*}{\Rightarrow} x$ *if and only if* x *contains equal number of* a*'s and* b*'s.*

(b) $A \overset{*}{\Rightarrow} x$ *if and only if* x *contains one more* a *than* b.

(c) $B \overset{*}{\Rightarrow} x$ *if and only if* x *contains one more* b *than* a.

Thus, we see that the start variable S of the given grammar G derives all strings with equal number of a's and b's and nothing else.

□