Some Context Free Languages
An Expostion
by
William Gasarch

1 Context Free Grammars and Languages

Definition 1.1 A *Context Free Grammar* is a tuple $G = (N, \Sigma, R, S)$ such that:

- N is a finite set of *nonterminals*
- Σ is a finite alphabet. Note $\Sigma \cap N = \emptyset$.
- $R \subseteq N \times (N \cup \Sigma)^*$.
- $S \in N$, the start symbol.

If S can generate $w \in (\Sigma \cup N)^*$ we denote this $S \Rightarrow w$.

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow w\}$$

Definition 1.2 A language L is a *Context Free Language* if there exists a context free grammar G such that $L(G) = L$.

In this document we will show several languages are context free.
We will need the following definition for some of the proofs.

2 $L = \{a^n b^n : n \in \mathsf{N}\}$ is a CFL

Here is the context free language G:
$S \rightarrow aSb$
$S \rightarrow e$
The proof that $L(G) = L$ is an easy induction on the number-of-steps in a derivation, which we omit.

3 $L = \{w : \#_a(w) = \#_b(w)\}$ is a CFL

Theorem 3.1 *Let the language L below is a CFL.*

$$L = \{w : \#_a(w) = \#_b(w)\}$$

Proof:

Let G the the following context free grammar.

$S \to aSb \quad | \quad bSa$

$S \to SS.$

$S \to e.$

We show that $L(G) = L$.

$L(G) \subseteq L$ is an easy induction on the number-of-steps in a derivation, which we omit.

We prove that, for all $w \in L$, $w \in L(G)$ by induction on $|w|$.

Base Case $|w| = 0$ so $w = e$. This is clearly in $L(G)$ using $S \to e$.

Ind Hyp Let $n \geq 1$. For all $w \in L$ of length$\leq n-1$, $w \in L(G)$.

Ind Step Let $w \in L$, $|w| = n$. We show $w \in L(G)$. We assume n is even.

Case 1 $w = aw'b$. Then $w' \in L$, $|w'| = n - 2len - 1$. By the IH, $S \Rightarrow w'$. Hence we have

$$S \to aSb \Rightarrow aw'b = w$$

Case 2 $w = bw'a$. Similar to case 1.

Case 3 $w = aw'a$. Let $w = w_1 w_2 \cdots w_n$ where $w_1 = w_n = a$.

For $1 \leq i \leq n$ let $x_i = w_1 \cdots w_i$ and $r_i = \frac{\#_b(w_i)}{\#_a(w_i)}$. Note $r_1 = 0$ and $r_{n-1} = \frac{n-1}{n}$

Claim There exists $2 \leq k \leq n-2$ such that $r_k = 1$.

Proof of Claim

Since $r_1 < 1$ and $r_{n-1} > 1$ there exists a least k, $2 \leq k \leq m-1$, such that $r_k \geq 1$. If $r_k = 1$ then we are done. So we assume $r_k > 1$. Since i is the least such we have $r_{k-1} < 1$. Hence

$r_{k-1} = \frac{\#_b(x_{k-1})}{\#_a(x_{k-1})} < 1$

$r_k = \frac{\#_b(x_k)}{\#_a(x_k)} > 1$

Since $r_{k-1} < r_k$, $w_k = b$. Hence $\#_b(x_{k-1}) = \#_b(x_k) - 1$ and $\#_a(x_k) = \#_a(x_k)$

Hence we have
$r_{k-1} = \frac{\#_b(x_k)-1}{\#_a(x_k)} < 1$
$r_k = \frac{\#_b(x_k)}{\#_a(x_k)} > 1$
The first equation yields

$$\#_b(x_k) - 1 < \#_a(x_k).$$

The second equation yields.

$$\#_b(x_k) > \#_a(x_k)$$

which we rewrite as

$$\#_a(x_k) < \#_b(x_k)$$

Combining the $<$ inequalities we get

$$\#_b(x_k) - 1 < \#_a(x_k) < \#_b(x_k).$$

Since all of the quantities are natural numbers this cannot occur. Hence the case where $r_k > 1$ cannot occur, so $r_k = 1$.

So we have $w = xy$ where $x, y \neq e$ and $\#_a(x) = \#_b(x)$, so $x \in L$. Since $w \in L$, we also have $\#_a(y) = \#_b(y)$, so $y \in L$. By the Induction Hypothesis $x, y \in L(G)$. Hence $S \Rightarrow y$ and $S \Rightarrow x$. Therefore $w \in L(G)$ as follows:

$$S \to SS \Rightarrow xy = w.$$

Case 4 $w = bw'b$. Similar to Case 3. ∎

4 A Useful Lemma

In the proof of Theorem 3.1, Case 3, we had to show that a string $w \in L$ that began with an a ended with a b must be of the form xy where $x \in L$ and $y \in L$. We prove a general lemma using the proof of that claim.

Lemma 4.1 *Let $m \in \mathsf{N}$. Let*

$$L_0 = \{w : \#_b(w) = m\#_a(w) + 0\}.$$

$$L_1 = \{w : \#_b(w) = m\#_a(w) + 1\}.$$

$$\vdots$$

$$L_{m-1} = \{w : \#_b(w) = m\#_a(w) + m - 1\}.$$

Let $w \in L_0$. Let $w = w_1 \cdots w_{(m+1)n}$. (There are n a's and mn b's.) For $1 \le k \le (m+1)n$ let $x_k = w_1 \cdots w_k$ and $r_k = \frac{\#_b(x_k)}{\#_a(x_k)}$.

1. *If there exists $1 \le i < j < (m+1)n$ such that $r_i < m$, $r_j > m$, and $\#_a(x_i) \ge 1$ then there exists k, $i < k < j$, where $r_k = m$. Hence there exists $x, y \ne e$ such that $w = xy$ and $x, y \in L_0$. (This follows since if $r_k = m$ then $x_k \in L_0$, so the rest of the string is also in L_0.)*

2. *If there exists $1 \le i < j < (m+1)n$ such that $r_i > m$, $r_j < m$, and $\#_a(x_j \cdots x_{(m+1)n}) \ge 1$, then there exists k, $i < k < j$, where $r_k = m$. Hence there exists $x, y \ne e$ such that $w = xy$ and $x, y \in L_0$. This can be obtained by applying Part 1 to w^R.*

3. *If w begins with an a and ends with a b then one of the following occurs.*
   (a) *$w = xy$ where $x, y \in L_0$.*
   (b) *$w = axb^m$ where $x \in L_0$.*

4. *If w begins with a b and ends with an a then one of the following occurs.*
   (a) *$w = xy$ where $x, y \in L_0$.*
   (b) *$w = b^m xa$ where $x \in L_0$.*

   This can be obtained by applying Part 3 to w^R.

5. *If w begins and ends with a b then either*
   (a) *$w = xy$ where $x, y \in L_0$.*

(b) *$w = b^k a x b^{m-k}$ for some $1 \le k \le m-1$.*

(c) *FILL IN LATER*

Proof:
1) Since $r_i < m$ and $r_j > m$ there exists a least k, $i < k < j$, such that $r_k \ge m$. If $r_k = m$ then we are done. So we assume $r_k > m$. Since k is the least such number we know $r_{k-1} < m$. Hence

$r_{k-1} = \frac{\#_b(x_{k-1})}{\#_a(x_{k-1})} < m$ (Note that $\#_a(x_{k-1}) \ge 1$.)

$r_k = \frac{\#_b(x_k)}{\#_a(x_k)} > m$

Since $r_{k-1} < r_k$, $w_k = b$. Hence $\#_b(x_{k-1}) = \#_b(x_k) - 1$ and $\#_a(x_k) = \#_a(x_k)$

Hence we have

$r_{k-1} = \frac{\#_b(x_k)-1}{\#_a(x_k)} < m$

$r_k = \frac{\#_b(x_k)}{\#_a(x_k)} > m$

The first equation yields

$$\#_b(x_k) - 1 < m\#_a(x_k).$$

The second equation yields.

$$\#_b(x_k) > m\#_a(x_k)$$

which we rewrite as

$$m\#_a(x_k) < \#_b(x_k)$$

Combining the $<$ inequalities we get

$$\#_b(x_k) - 1 < m\#_a(x_k) < \#_b(x_k).$$

Since all of the quantities are natural numbers and $\#_a(x_k) \ge 1$ this cannot occur. Hence the case where $r_k > m$ cannot occur, so $r_k = m$.

3) w begins with an a and ends with a b. Let $i \ge 1$ be such that $w = aw'ab^i$. (The enumerated list here does not correlate with the one in the theorem; however, we always get one of the cases.)

1. If $1 \le i \le m-1$ then we will be applying Part 1 to the prefix a and the prefix aw'. The first ratio we need is $\frac{\#_b(a)}{\#_a(a)} = 0 < m$. The second ratio we need is

$$\frac{\#_b(aw')}{\#_a(aw')} = \frac{\#_b(w) - \#_b(ab^i)}{\#_a(w) - \#_a(ab^i)} = \frac{mn-i}{n-1} > m.$$

Hence Part 1 applies and we get $w = xy$ where $x, y \in L_0$.

2. If $i = m$ then the suffice $y = ab^i \in L_0$, so the prefix $x = aw' \in L_0$.

3. If $i \geq m+1$ then $w = aw'b^i = aw'b^{i-m}b^m$. Let $x = w'b^{i-m}$ and note that $w = axb^m$ and $x \in L_0$.

5) w begins with a b and ends with a b. Let $k, \ell \geq 1$ be such that $w = b^k aw'ab^\ell$. (The enumerated list here does not correlate with the one in the theorem; however, we always get one of the cases.)

1. $k \leq m-1$ and $\ell \leq m-1$. We apply Part 1 with $x_i = b^k a$ and $x_j = b^k aw'$. We have $\#_a(x_i) \geq 1$. we need $r_i < m$ and $r^j > m$.

   $\#_b(b^k a) = i$ and $\#_a(b^k a) = 1$ so $\frac{\#_b(b^k a)}{\#_a(b^k a)} = k < m$.

   $\#_b(b^k aw') = \#_b(w) - \#_b(ab^\ell) = mn - \ell$ and $\#_a(b^k aw') = \#_a(w) - \#_a(ab^\ell) = n-1$, so

   $\frac{\#_b(b^k aw')}{\#_a(b^k w')} = \frac{mn-\ell}{n-1} > m$.
   So Part 1 applies and $w = xy$ with $x, y \in L_0$.

2. $k = m$ or $\ell = m$. If $k = m$ then $w = b^m aw''$ so just take $x = b^m a$ and $y = w''$. Since $x \in L_0$, $y \in L_0$. The case of $\ell = m$ is similar.

3. $k \geq m+1$ and $\ell \geq m+1$.

4. $k \leq m-1$ or $\ell \geq m+1$. So $w = b^k aw'ab^{\ell+k-m}b^{m-k}$. Let $x = w'a$. Then $w = b^k x b^{m-k}$.

∎

5 $L = \{w : m\#_a(w) = \#_b(w)\}$ is a CFL

Theorem 5.1 *Let $m \ge 1$. The language L below is a CFL.*

$$L = \{w : m\#_a(w) = \#_b(w)\}$$

Proof:

Let G the the following context free grammar.
For every $\sigma_1 \cdots \sigma_{m+1}$ where m of the symbols are b and one of the symbols is a, and for every $0 \le i \le m+1$ we have the production
$S \rightarrow \sigma_1 \cdots \sigma_i S \sigma_{i+1} \cdots \sigma_{m+1}$.
$S \rightarrow SS$.
$S \rightarrow e$.
$S \rightarrow TaT$.
$T \rightarrow bS \quad | \; ST$.

1) $L(G) \subseteq L$.

We show by induction no the number-of-steps in a derivation that, for all $w \in \{a, b, S, T\}^*$ that G generates,

$$m(\#_a(w) + \#_T(w)) = \#_b(w).$$

Base Case If there is only one step them $w = e$ so the conclusion holds.
Ind Hyp If $w' \in \{a, b, S, T\}^*$ is generates by $n-1$ steps then

$$m(\#_a(w') + \#_T(w')) = \#_b(w').$$

Ind Step Let $S \Rightarrow w$ in n steps. Then $S \Rightarrow w'$ in $n-1$ steps and then some rule R goes from w' to w. By the IH.

$$m(\#_a(w') + \#_T(w')) = \#_b(w').$$

If R replaces an S with one a and m $b's$ then
$\#_a(w) = \#_a(w') + 1$.
$\#_b(w) = \#_b(w') + m$.
$\#_S(w) = \#_S(w')$.
$\#_T(w) = \#_T(w')$.
Hence

$$m((\#_a(w) - 1) + \#_T(w) = \#_b(w) - m$$

$$m\#_a(w) - m + \#_T(w) = \#_b(w) - m$$

$$m\#_a(w) + \#_T(w) = \#_b(w)$$

BILL - DO THE REST LATER

is an easy induction on the number-of-steps in a derivation, which we omit.

We prove that, for all $w \in L$, $w \in L(G)$ by induction on $|w|$.

Base Case $|w| = 0$ so $w = e$. This is clearly in $L(G)$ using $S \rightarrow e$.

Ind Hyp Let $n \geq 1$. For all $w \in L$ of length$\leq n-1$, $w \in L(G)$.

Ind Step Let $w \in L$, $|w| = n$. We show $w \in L(G)$.

Case 1 $w = aw'a$.

1. For the prefix a we have $\frac{\#_b(a)}{\#_a(a)} = 0 < m$. Also note that $\#_a(a) = 1 \geq 1$.

2. For the prefix aw' we have $\frac{\#_b(a)}{\#_a(a)} = \frac{\#_b(w)}{\#_a(w)-1} > \frac{\#_b(w)}{\#_a(w)} = m$.

By Lemma 4.1, there exists $x, y \in L$ such that $w = xy$. By the Induction Hypothesis $S \Rightarrow x$ and $S \Rightarrow y$. Hence

Case 2 $w = aw'b$. Let i be such that $w = aw''ab^i$.

Case 2.1 $0 \leq i \leq m-1$.

1. For the prefix a we have $\frac{\#_b(a)}{\#_a(a)} = 0 < m$

2. For the prefix $aw''a$ we have $\frac{\#_b(a)}{\#_a(a)} = \frac{\#_b(w)-i}{\#_a(w)} > \frac{\#_b(w)}{\#_a(w)} = m$.

By Lemma 4.1, there exists $x, y \in L$ such that $w = xy$. By the Induction Hypothesis

$S \Rightarrow x$ and $S \Rightarrow y$. Hence

$$S \rightarrow SS \Rightarrow xy = w$$

Case 2.2 $i \geq m$. So $w = aw'''ab^{i-m}b^m$. Note that $w = w'''ab^{i-m} \in L$. By the induction hypothesis

$S \Rightarrow w'''ab^{i-m}$. Hence

$$S \to aSb^m \Rightarrow aw'''ab^{i-m}b^m = aw'''b^i = w$$

Case 3 $w = bw'a$. Similar to Case 3.
Case 4 $w = bw'b$. We cannot use Lemma 4.1 with the prefix b since $\#_a(b) = 0$. We need to find the first a. Let i be such that $w = b^i aw''b$.
Case 4.1 $i \le m - 1$.

1. For the prefix $b^i a$ we have $\frac{\#_b(b^i a)}{\#_a(b^i a)} = \frac{i}{1} = i < m$. Note that $\#_a(b^i a) = 1 \ge 1$.
2. For the prefix $aw''a$ we have $\frac{\#_b(a)}{\#_a(a)} = \frac{\#_b(w)-i}{\#_a(w)} > \frac{\#_b(w)}{\#_a(w)} = m$.

By Lemma 4.1, there exists $x, y \in L$ such that $w = xy$. By the Induction Hypothesis
$S \Rightarrow x$ and $S \Rightarrow y$. Hence

∎