CS 381

Supplement to Reba's lecture: Computing $L(G)$ for CFG G

10/15/01

First, let me give you a nice clean version of the proof I did in class:

Let $G = (\Sigma, N, P, S)$, where $\Sigma = \{a, b\}, N = \{S\}, P = \{S \rightarrow \varepsilon, S \rightarrow aSb\}$

Claim: $L(G) = \{a^n b^n \mid n \in \mathbb{N}\}$

Proof:

$\{a^n b^n \mid n \in \mathbb{N}\} \subseteq L(G)$**:** We proceed by induction on n. Base case: if n = 0, $a^n b^n = \varepsilon \in L(G)$ since $S \rightarrow \varepsilon$ is a production in P. Inductive step: Suppose $a^n b^n \in L(G)$. Then there exists a derivation $S \overset{*}{\underset{G}{\rightarrow}} a^n b^n$. Now, this gives us the derivation $S \underset{G}{\rightarrow} aSb \overset{*}{\underset{G}{\rightarrow}} a(a^n b^n)b = a^{n+1}b^{n+1}$, where the first arrow is via the production $S \rightarrow aSb$, and the second is via the derivation that must exist by the inductive hypothesis.

$L(G) \subseteq \{a^n b^n \mid n \in \mathbb{N}\}$**:** For $x \in L(G)$, we proceed by induction on the length of the G-derivation of x. Base case: if $S \underset{G}{\rightarrow} x$, then $x = \varepsilon = a^0 b^0$. Inductive step: Assume that if $S \overset{m}{\underset{G}{\rightarrow}}$ then $x = a^n B^n$ for some $n \in \mathbb{N}$. Now suppose $S \overset{m+1}{\underset{G}{\longrightarrow}} x$. This derivation must begin with the production $S \rightarrow aSb$, so it has the form $S \overset{1}{\underset{G}{\rightarrow}} aSb \overset{m}{\underset{G}{\rightarrow}}$. But then $x = ayb$ for some $y \in \Sigma^*$ such that $S \overset{m}{\underset{G}{\rightarrow}} y$. Now, by the inductive hypothesis, $y = a^n b^n$ for some $n \in \mathbb{N}$, so $x = a(a^n b^n)b = a^{n+1}b^{n+1}$ for that n.

Here's another, more difficult example, taken from *Introduction to Automata Theory, Languages, and Computation* by Hopcroft and Ullman . Let $G = (\Sigma, N, P, S\}$, where $\Sigma = \{a, b\}$, $N = \{S, A, B\}$, and $P = \{S \rightarrow aB,\ S \rightarrow bA,\ A \rightarrow a,\ A \rightarrow aS,\ A \rightarrow bAA,\ B \rightarrow b,\ B \rightarrow bS,\ B \rightarrow aBB\}$.

Claim: $L(G) = \{w \in \{a, b\}^+ \mid \#_a(w) = \#_b(w)\}$

Proof:

Inductive Hypothesis: For $w \in \{a, b\}^+$,

1. $S \overset{*}{\underset{G}{\to}} w$ if and only if w contains an equal number of a's and b's.

2. $A \overset{*}{\underset{G}{\to}} w$ if and only if w has one more a than it has b's.

3. $B \overset{*}{\underset{G}{\to}} w$ if and only if w has one more b than it has a's.

We proceed by induction on $|w|$. Base case: If $|w| = 1$, then either w = a, or w =b. Since no string of length 1 is derivable from S, part 1 of the inductive hypotheses holds. Part 2 holds because the production $A \to a$ is in P, and because this production and $B \to b$ are the only ones that don't increase the length of the string to which they are applied (thus, a is the only string of length 1 derivable from A). Similarly, part 3 holds.

Inductive step. Assume that the inductive hypothesis holds for all w such that $|w| \leq k - 1$. We show that part 1 of the induction hypothesis holds for $|w| = k$. (Showing parts 2 and 3 is similar and left to the reader.)

Suppose $|w| = k$, and $S \overset{*}{\underset{G}{\to}} w$. We must show that w contains an equal number of a's and b's. Now, the derivation must begin with either $S \overset{*}{\underset{G}{\to}} aB$ or $S \overset{*}{\underset{G}{\to}} bA$. In the former case, w has the form aw_1, where $|w_1| = k - 1$, and $B \overset{*}{\underset{G}{\to}} w_1$. By the inductive hypothesis, the number of b's in w_1 is one more than the number of a's, so w has an equal number of a's and b's. The latter case is analogous.

Now, suppose $|w| = k$, and w has an equal number of a's and b's. We must show that $w \in L(G)$. Either the first letter of w is an a, or it is a b. Assume $w = aw_1$. Then $|w_1| = k - 1$, and w_1 has one more b than it has a's, so by the inductive hypothesis, $B \overset{*}{\underset{G}{\to}} w_1$. Thus, we have a derivation $S \overset{*}{\underset{G}{\to}} aB \overset{*}{\underset{G}{\to}} aw_1 = w$. If, instead, the first letter of w is b, the argument is analogous.