Context-free languages II

Thursday, February 18, 2021 11:45 AM

Closure properties:

(1) If L_1, L_2 are CFL then $L_1 \cup L_2$ is a CFL

(2) If L_1, L_2 are CFL Then so is $L_1 \circ L_2$.

(3) If L is a CFL then so is L^*

PROOFS

(1) $G_1 = (V_1, T_1, S_1, \cdots)$ $L_1 = L(G_1)$

$G_2 = (V_2, T_2, S_2, \cdots)$ $L_2 = L(G_2)$

For $L_1 \cup L_2$ we just take the unions of T_1 and T_2, V_1 and V_2, new start symbol S & add the rules

$S \to S_1 \mid S_2$

same thing as $\begin{cases} S \to S_1 \\ S \to S_2 \end{cases}$

(2) $S \to S_1 S_2$

(3) New start symbol S' and new rules

$S' \to S S' \mid \varepsilon$

NON-RESULTS

(i) the complement of a CFL may not be a CFL.

(ii) The intersection of 2 CFLs may not be a CFL.

If L is a CFL and R is a regular language then $L \cap R$ is context free.

EXAMPLES of CFG design:

Two techniques: (i) use recursion

(ii) use matching

(1) $\Sigma = \{a, b\}$

$L = \{a^n b^{2n} \mid n \geq 0\}$

Each a is matched with 2 b's.

$S \to aSbb \mid \varepsilon$

(2) $L = \{x \in \Sigma^* \mid x = x^{REV}\}$

x^{REV} means x written backwards

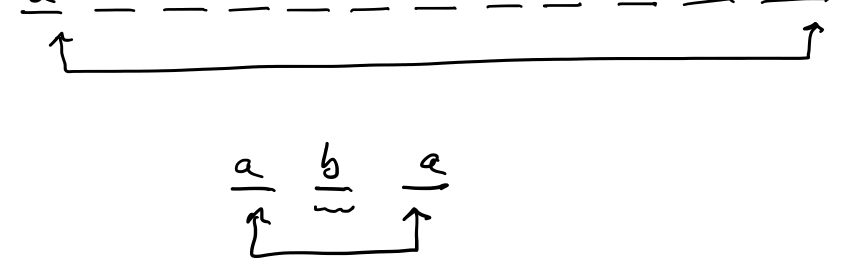
$(abaa)^{REV} = aaba \notin L$

$(abba)^{REV} = abba \in L$

PALINDROME

MADAM I'M ADAM

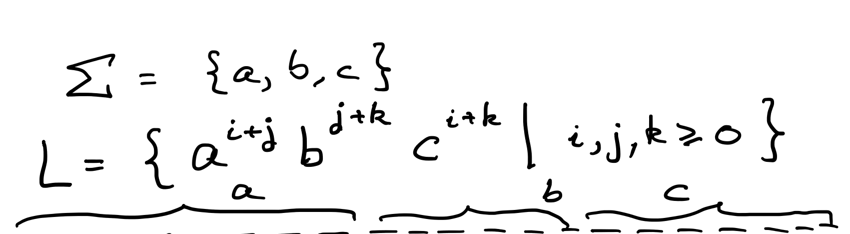
MALAYALAM



$S \to aSa \mid bSb \mid a \mid b \mid \varepsilon$

(3) $\Sigma = \{a, b, c\}$

$L = \{a^{i+j} b^{j+k} c^{i+k} \mid i, j, k \geq 0\}$



$S \to aSc \mid AB \mid \varepsilon$

$A \to aAb \mid \varepsilon$

$B \to bBc \mid \varepsilon$

EXAMPLES based on find self-similar structures

(4) $\Sigma = \{a, b\}$

$L = \{x \in \Sigma^* \mid$ each prefix of x has at least as many a's as b's$\}$

aaabab ✓

)a)a)b)b)b)aa ✗

$x \in L$ could be ε

If $x \in L$ and $x \neq \varepsilon$ what is the first letter of x?

It has to be an a

So $x = ay$

Now y may or may not be in L.

- If y is in L we can handle this case with $S \to aS$
- If y is not in L, some prefix of y has more b's than a's.

Let u be the shortest prefix with this property. u will have 1 more b than # of a's. So $u = wb$ so $y = wbv$

Now $x = awb|v$ is in L so

awb has as many a's as b's

v must satisfy the property as well.

Thus w, v are both in L.

$S \to aSbS$

$S \to \varepsilon \mid aS \mid aSbS$

---

(5) $L = \{x \in \Sigma^* \mid x$ has equal # of a's and b's$\}$

e.g. $bbbaaaab \in L$

$aba \notin L$

$d(x) := \#_b(x) - \#_a(x)$

$L = \{x \mid d(x) = 0\}$

suppose $x \in L$, $x \neq \varepsilon$ non-empty

Let u be the shortest prefix of x s.t. $u \in L$

e.g. $\underbrace{\underbrace{bbaa}_{u}ab}_{x}$

$d(u) = 0$

Suppose u starts with b, then it must end with a. [WHY?]

$u = b\underline{v}a$

↳ this must be balanced so $d(v) = 0$

$x = uz$

if $d(x) = 0$ and $d(u) = 0$

then $d(z) = 0$

$x = a\underline{v}b\underline{z}$ $v, z \in L$

$S \to aSbS$

similarly we reason in the opposite case $S \to bSaS$

our complete grammar is

$S \to aSbS \mid bSaS \mid \varepsilon$

another valid but different grammar for the same language:

$S \to aSb \mid bSa \mid SS \mid \varepsilon$

ALG. Given a CFG G how do I know if it generates anything?

Given G is $L(G) = \emptyset$?

Given a CFG we say $X \in V$ is generating if $X \xrightarrow{*} w \in \Sigma^*$

$\xrightarrow{*}$ produces after possibly several steps.

FACT $L(G) \neq \emptyset$ iff S is generating.

We will consider all terminal symbols to be generating.

GEN: set of all generating symbols.

Put all terminal symbols in GEN.

Do until GEN does not change anymore:

{ For each rule $X \to \alpha$ verify if every symbol in α is in GEN already. If so add X to GEN.

Check if $S \in GEN$.

EXAMPLE

$S \to AB \mid a \quad A \to b$

$GEN_0 = \{a, b\}$

$GEN_1 = \{S, a, b, A\}$ STOP.

YES $L(G) \neq \emptyset$

$B \notin GEN$ so any rules with B in the RHS should be removed.

$S \to a \quad \cancel{A \to b}$

Hey A is unreachable.

$S \to a$

$L(G) = \{a\}$

If $X \to \varepsilon$ you mark $X \in GEN$.