CS3133

HW#4

DUE: Tuesday, September 30

1. Give a context-free grammar for each of the following languages.

a (3 points) $\left\{01(0011)^n 11(10)^n \mid n \geq 0\right\}$.

b (3 points) $\left\{0^m 1^n 2^{m+n} \mid m, n \geq 0\right\}$.

c (3 points) $\left\{0^n w w^r 1^n \mid n \geq 0 \wedge w \in \{0,1\}^*\right\}$ where w^r is *w* written in reverse order. For example, $(011)^r = 110$.

2. (10 points) For CFG $G = (N, \Sigma, P, S)$, we call $x \in N \cup \Sigma$ *untouchable* if there does not exist a derivation $S \overset{*}{\underset{G}{\longrightarrow}} \alpha x \beta$ for some $\alpha, \beta \in (N \cup \Sigma)^*$. Prove or give a counterexample to the following.

   CONJECTURE: For any context free language *L* there exists a CFG grammar *G* without untouchable symbols such that $L = L(G)$.

3. ***a*** (8 points) Give a grammar equivalent to

$$S \to ASA \mid aB$$
$$A \to B \mid S$$
$$B \to b \mid \varepsilon$$

which does not have any unit productions or $\varepsilon-$productions.

b (4 points) Give a grammar equivalent to

$$S \to ASA \mid aB$$
$$A \to B \mid S$$
$$B \to b \mid \varepsilon$$

in Chomsky Normal Form.

4. (10 points) For any language *L* we define a new language, $Suff(L)$, which is the set of all strings which are suffixes of strings of *L*. That is,

$$Suff(L) = \left\{x_j ..., x_n \mid 1 \leq j \leq n \wedge x_1 ... x_n \in L\right\} \cup \{\varepsilon\} \quad .$$

So if $L = \{001, 11\}$, then $Suff(L) = \{001, 01, 1, \varepsilon, 11\}$. Prove or give a counterexample to the following

Conjecture: The set of all context-free languages is closed under the *Suff* operation. That is, for any context-free language *L*, the language $Suff(L)$ is also context-free.

CS3133

Solutions for HW#4

1. ***a***

$S \to 01A$

$A \to 0011A10|11$

b

$S \to 0S2|A$

$A \to 1A2|\varepsilon$

c

$S \to 0S1|A$

$A \to 0A0|1A1|\varepsilon$

2. If L is a CFL, then there exists CFG $G = (N, \Sigma, P, S)$ such that $L(G) = L$. Clearly, we can remove all untouchable symbols from $N \cup \Sigma$, and all productions containing these symbols, without changing the language that G generates. We proceed by identifying *touchable* symbols, τ, and then the untouchable symbols are $(N \cup \Sigma)/\tau$.

$\tau \leftarrow \{S\}$

for each $A \in N \cap \tau$ **do**

  for each production $(A \to \alpha) \in P$

    add each symbol of α to τ

until τ doesn't change

remove from P all productions containing a symbol not in τ

3. ***a***A and B are nullable. Removing $\varepsilon-$productions yields

$$S \to ASA|SA|AS|S|aB|a$$

$$A \to B|S$$

$$B \to b$$

To remove unit productions, we first compute *CHAIN*(S)={S}, *CHAIN*(A)={B,S}, *CHAIN*(B)=$\varnothing$. Then we remove the unit productions

$$S \to ASA|SA|AS|aB|a$$

$$A \to b|ASA|SA|AS|aB|a$$

$$B \to b$$

b

$$S \to AB_1|SA|AS|B_aB|a$$

$$B_1 \to SA$$

$$B_a \to a$$

$$A \to b|AB_1|SA|AS|B_aB|a$$

$$B \to b$$

5. The CONJECTURE is true. If L is a context-free language, there must be a CFG in Chomsky Normal Form $G=(N,\Sigma,P,S)$ such that $L=L(G)-\{\varepsilon\}$. We construct $G^*=(N^*,\Sigma,P^*,S_{suff})$ from G, satisfying $Suff(G)=L(G^*)$, in the following way. For each $A\in N$ we add A_{suff} to the nonterminals. That is, $N^*=N\cup\{A_{suff}\,|\,A\in N\}$. We note that G^* is not in CNF.

We do not change any of the productions $A\rightarrow w$ for any $A\in N$, so the string of terminals which can be derived from each $A\rightarrow w$ is unchanged. That is, the language generated from (N,Σ,P,A) is the same as the language generated from (N^*,Σ,P^*,A). For each production of the form $A\rightarrow a$ in P, $A\in N$ and $a\in\Sigma$, we add the productions $A_{suff}\rightarrow a\,|\,\varepsilon$ to P^*. For each production of the form $A\rightarrow BC$ in P, $A,B,C\in N$, we add the productions $A_{suff}\rightarrow AB|A_{suff}B|B|B_{suff}$ to P^*.