EECS 2001N : Introduction to the Theory of Computation

Suprakash Datta
Office: LAS 3043

Course page: http://www.eecs.yorku.ca/course/2001N
Also on Moodle

Simpler CFG's: Chomsky Normal Form

A CFG $G = (V, \Sigma, R, S)$ is in "Chomsky normal form" if every rule is of the form

- $A \to BC$, $A, B, C \in V$, or
- $A \to x$,
  with variables $A \in V$ and $B, C \in V \setminus \{S\}$, and $x \in \Sigma$
  Note: this implies that the start variable cannot be on the right side of a rule
- For the start variable S we also allow the rule $S \to \epsilon$
- Obvious fact: All regular languages are also context free languages

Advantage: Grammars in this form are far easier to analyze

Converting a CFG to Chomsky Normal Form

Outline of Proof:

- We rewrite every CFG in Chomsky normal form
- We do this by replacing, one-by-one, every rule that is not 'Chomsky'
- We have to take care of:
  - Starting Symbol
  - ϵ symbol,
  - all other violating rules.

Converting a CFG to CNF - details (Pg 104)

Given a context-free grammar $G = (V, \Sigma, R, S)$, rewrite it to Chomsky Normal Form by

- New start symbol S_0 (and add rule $S_0 \rightarrow S$)
- Remove $A \rightarrow \epsilon$ rules (*from the tail*):
  before: $B \rightarrow xAy$ and $A \rightarrow \epsilon$, after: $B \rightarrow xAy|xy$
- Remove unit rules $A \rightarrow B$ (*by the head*):
  before: $A \rightarrow B$ and $B \rightarrow xCy$ after: $A \rightarrow xCy$ and $B \rightarrow xCy$
- Shorten all rules to two:
  before: $A \rightarrow B_1 B_2 \ldots B_k$, after: $A \rightarrow B_1 A_1$, $A_1 \rightarrow B_2 A_2$, $\ldots$, $A_{k-2} \rightarrow B_{k-1} B_k$
- Replace ill-placed terminals a by T_a with $T_a \rightarrow a$

Converting a CFG to CNF - points to note

- Do not re-introduce rules removed earlier
- Example: $A \rightarrow A$ simply disappears
- When removing $A \rightarrow \epsilon$ rules, insert **all** new replacements: $B \rightarrow AaA$ becomes $B \rightarrow AaA|aA|Aa|a$

Converting a CFG to CNF - Example

Initial CFG:

$$S \to aSb|\epsilon$$

CNF:

- $S_0 \to \epsilon|T_aT_b|T_aX$
- $X \to ST_b$
- $S \to T_aT_b|T_aX$
- $T_a \to a$
- $T_b \to b$