Chomsky Normal Form, Conversion (CFG-CNF)

Chomsky Normal Form (CNF)

⇒ In Chomsky Normal Form (CNF), We have a restriction on the length of Right Hand Side (R.H.S) where elements in RHS should either be two variables or a terminal.

⇒ A CFG is in CNF if the productions are in the following forms:

$A \rightarrow a$
$A \rightarrow BC$

where A, B and $C \in V$,
$a \in T$

CNF

⇒ Steps to Convert a given CFG to CNF Form:

Step1: If the Start Symbol S occurs on some right side of production rules then Create a new Start Symbol S' and add a new production $S' \to S$, otherwise add the Start Symbol S in production rules.

Step2: Remove NULL production. ✓

Step3: Remove UNIT production. ✓

Step4: Replace each production $A \to B_1 \ldots B_n$ where $n > 2$, with $A \to B_1 C$ where $C \to B_2 \ldots B_n$.
Repeat this step for all productions having two or more Symbols on the right side.

Step5: If any production is in the form of $A \to aB$, where $a \in T$ and $A, B \in V$ then the production is replaced by $A \to XB$ and $X \to a$.

Repeat this step for every production which is in the form of $A \to aB$.

Conversion of CFG to CNF

⇒ Convert the following CFG to CNF:

$P: S \to ASA \mid aB, \; A \to B \mid S, \; B \to b \mid \epsilon$

Step 1: $P: S' \to S, \; S \to ASA \mid aB, \; A \to B \mid S, \; B \to b \mid \epsilon$

Step 2: Remove the null productions: $B \to \epsilon$ and $A \to \epsilon$

After Removing $B \to \epsilon$:

$P: S' \to S, \; S \to ASA \mid aB \mid a, \; A \to B \mid S \mid \epsilon, \; B \to b$

After Removing $A \to \epsilon$:

$P: S' \to S, \; S \to ASA \mid aB \mid a \mid AS \mid SA \mid S,$
$A \to B \mid S, \; B \to b$

Step 3: Remove the Unit Productions: $S \to S$, $S' \to S$, $A \to B$ and $A \to S$

After Removing $S \to S$:

$P: S' \to S, \; S \to ASA \mid aB \mid a \mid AS \mid SA, \; A \to B \mid S,$
$B \to b$

After Removing $S' \to S$: $P:$

$S' \to ASA \mid aB \mid a \mid AS \mid SA$

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

$A \to B \mid S$

$B \to b$

After Removing $A \to B$: P: $S' \to ASA \mid aB \mid a \mid AS \mid SA$,

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

$A \to b \mid S$, $B \to b$.

After Removing $A \to S$: P: $S' \to ASA \mid aB \mid a \mid AS \mid SA$,

$S \to ASA \mid aB \mid a \mid AS \mid SA$,

$A \to b \mid ASA \mid aB \mid a \mid AS \mid SA$,

$B \to b$

Step 4: Now find out the productions that has more than two variables in RHS.

$S' \to ASA$, $S \to ASA$, $A \to ASA$

After Removing: P: $S' \to A\underline{X} \mid aB \mid a \mid AS \mid SA$,

$S \to A\underline{X} \mid aB \mid a \mid AS \mid SA$,

$A \to b \mid A\underline{X} \mid aB \mid a \mid AS \mid SA$.

$B \to b$

$\underline{X \to SA}$

$S' \to aB$, $S \to aB$, $A \to aB$

P: $S' \to AX \mid YB \mid a \mid AS \mid SA$,

$S \to AX \mid YB \mid a \mid AS \mid SA$,

$A \to b \mid AX \mid YB \mid a \mid AS \mid SA$;

$B \to b$,

$X \to SA$,

$Y \to a$

CNF