A CFG is in Chomsky Normal Form if the Productions are in the following forms −

* A → a
* A → BC
* S → ε

where A, B, and C are non-terminals and **a** is terminal.

Algorithm to Convert into Chomsky Normal Form −

**Step 1** − If the start symbol **S** occurs on some right side, create a new start symbol **S’** and a new production **S’→ S**.

**Step 2** − Remove Null productions. (Using the Null production removal algorithm discussed earlier)

**Step 3** − Remove unit productions. (Using the Unit production removal algorithm discussed earlier)

**Step 4** − Replace each production **A → B1…Bn** where **n > 2** with **A → B1C** where **C → B2 …Bn**. Repeat this step for all productions having two or more symbols in the right side.

**Step 5** − If the right side of any production is in the form **A → aB** where a is a terminal and **A, B** are non-terminal, then the production is replaced by **A → XB** and **X → a**. Repeat this step for every production which is in the form **A → aB**.

Problem

Convert the following CFG into CNF

S → ASA | aB, A → B | S, B → b | ε

Solution

**(1)** Since **S** appears in R.H.S, we add a new state **S0** and **S0→S** is added to the production set and it becomes −

S0→S, S→ ASA | aB, A → B | S, B → b | ∈

**(2)** Now we will remove the null productions −

B → ∈ and A → ∈

After removing B → ε, the production set becomes −

S0→S, S→ ASA | aB | a, A → B | S | ∈, B → b

After removing A → ∈, the production set becomes −

S0→S, S→ ASA | aB | a | AS | SA | S, A → B | S, B → b

**(3)** Now we will remove the unit productions.

After removing S → S, the production set becomes −

S0→S, S→ ASA | aB | a | AS | SA, A → B | S, B → b

After removing S0→ S, the production set becomes −

S0→ ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → B | S, B → b

After removing A→ B, the production set becomes −

S0 → ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → S | b

B → b

After removing A→ S, the production set becomes −

S0 → ASA | aB | a | AS | SA, S→ ASA | aB | a | AS | SA

A → b |ASA | aB | a | AS | SA, B → b

**(4)** Now we will find out more than two variables in the R.H.S

Here, S0→ ASA, S → ASA, A→ ASA violates two Non-terminals in R.H.S.

Hence we will apply step 4 and step 5 to get the following final production set which is in CNF −

S0→ AX | aB | a | AS | SA

S→ AX | aB | a | AS | SA

A → b |AX | aB | a | AS | SA

B → b

X → SA

**(5)** We have to change the productions S0→ aB, S→ aB, A→ aB

And the final production set becomes −

S0→ AX | YB | a | AS | SA

S→ AX | YB | a | AS | SA

A → b A → b |AX | YB | a | AS | SA

B → b

X → SA

Y → a