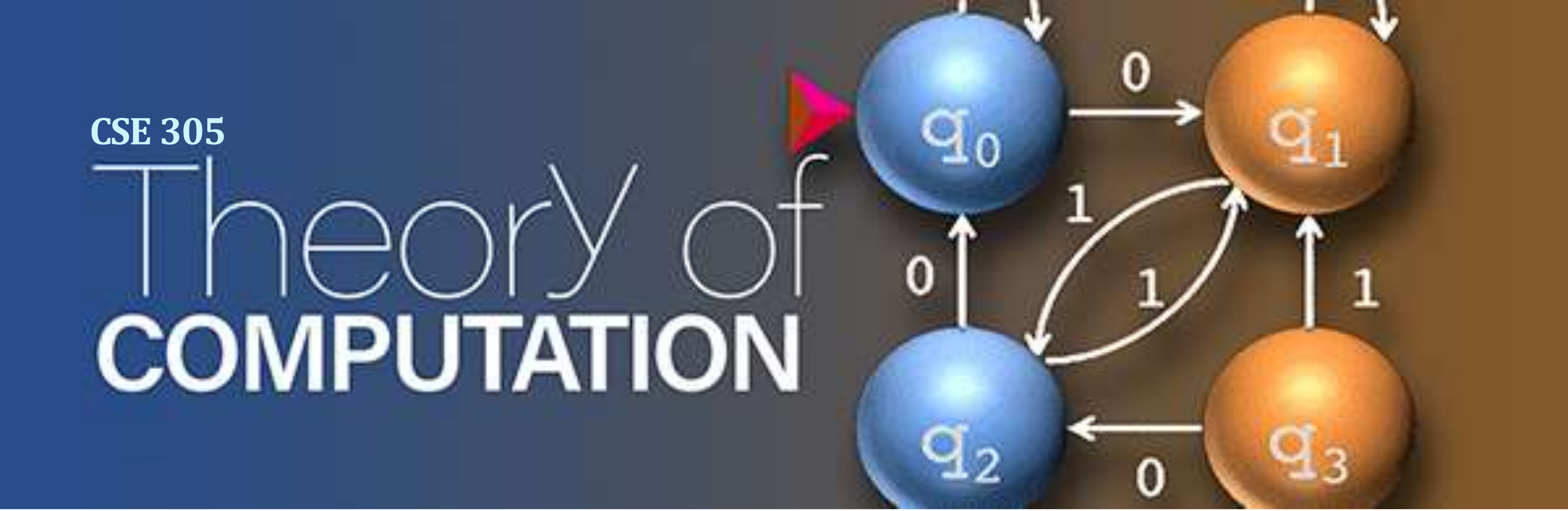

Lecture 25

Context-Free Grammars (3)

Md. Mijanur Rahman, Prof. Dr.

Dept. of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Bangladesh.

www.mijanrahman.com

Contents

Context-Free Grammars

Simplification of CFG

- **To simplify a given grammar G, we try to eliminate the symbols and the productions in the grammar which do not take part in derivation of sentences.**
- The simplification does not reduce the language generation power of the CFG G, that is, it will generate the same language after simplification as G generates.
- So, the purpose of simplifying the CFG is to eliminate all such productions that do not participate in derivation of a string in the language.

Simplification of CFG

- The following symbols and productions are removed from the CFG:
  1. We must eliminate all symbols which produce empty (ε) directly or indirectly. All such symbols are called **nullable symbols or nullable non-terminals or nullable variables**.
  2. We must eliminate all productions of the form $X \rightarrow Y$. **These productions are called unit productions**.
  3. We must eliminate all symbols that do not appear in derivation of any string from the start symbol. **Such symbols are called useless symbols and all such productions are called useless productions.**
- **A symbol of CFG is called useful symbol if it appears in the derivation of at least one string of the language generated by that CFG.**

Removal of Null/Empty (ε) Production

- **Production of the form X ⇒ ε arc called empty or null productions.** Thus, we can define null productions as any production in the grammar which generates ε directly (in a single step) or indirectly (in many steps).
- **It is also called ε (or empty) production and the non-terminals which lead to ε (empty) are called nullable non-terminals or variables.**
- If ε is an element of the language generated by the grammar then we cannot remove ε; otherwise we can remove ε productions.

Removal of Null/Empty (ε) Production

- **Procedure for Removing Null Production:**
- The steps that need to be followed for removal of ε productions are listed below:
  - **Step 1.** Identify all the nullable non-terminals.
  - **Step 2.** Remove all the productions which directly produce ε (empty).
  - **Step 3.** If there is a production of the form $X \rightarrow \alpha$ and α contains nullable non-terminals then add all new productions which are obtained after deleting the subset of the nullable non-terminals from α.

Removal of Null/Empty (ε) Production

Example: Remove the empty or null productions from the following grammars:

$$E \longrightarrow AXA$$
$$A \longrightarrow aA \mid \epsilon$$
$$X \longrightarrow bX \mid \epsilon$$

Solution:

- Applying the procedure as described earlier to remove null productions:

- Identify all the nullable non-terminals.

Here, we can see that X and A are directly producing ϵ (i.e., $A \rightarrow \epsilon$ and $X \rightarrow \epsilon$) and when we place ϵ for A and X in Production 1 then E also produces (indirectly) ϵ . ($E \rightarrow AXA \rightarrow \epsilon\, XA \rightarrow \epsilon\epsilon\, A \rightarrow \epsilon\epsilon\epsilon \rightarrow \epsilon$) So, A, X, and E are nullable non-terminals.

Removal of Null/Empty (ε) Production

$E \longrightarrow AXA$

$A \longrightarrow aA \mid \epsilon$

$X \longrightarrow bX \mid \epsilon$

- Remove all productions which produce ϵ directly.

$A \longrightarrow \epsilon$ and $X \longrightarrow \epsilon$ are two productions in the given grammar. On removing these two productions the grammar becomes

$$E \longrightarrow AXA$$

$$A \longrightarrow aA \ (A \rightarrow \epsilon \text{ is removed})$$

$$X \longrightarrow bX \ (X \rightarrow \epsilon \text{ is removed})$$

- We replace the subset of nullable non-terminal by ϵ in RHS of the production. So, when we first replace A from AXA it will become XA. Similarly, when we replace both As it will become X and so on. The final answer would be written as

$$E \longrightarrow AXA \mid AX \mid XA \mid AA \mid X \mid A$$

$$A \longrightarrow aA \mid a$$

$$X \longrightarrow bX \mid b$$

Removal of Unit Production

- **Any production of the form A → B is called a unit production.** Thus, a unit production may be defined as a production which contains single non-terminals on both sides.
- Unit productions can be removed without affecting the language generation power of the grammar.

- **Procedure:**
  - **RHS non-terminal of the unit production can be replaced by its corresponding production in the grammar if any is given; otherwise we can drop such productions.**

Removal of Unit Production

$$E \longrightarrow E * T \mid T$$
$$T \longrightarrow E + E \mid (E) \mid id$$
$$T \longrightarrow T + F \mid F$$
$$F \longrightarrow (T) \mid id$$

Example: Remove unit productions from the grammar given below-

Solution:

- There are two unit productions:

$$T \longrightarrow F$$
$$E \longrightarrow T$$

- So, we can replace F from the RHS of the unit production by-

$$F \longrightarrow (T) \mid id$$

- So, the Production-3 becomes-

$$T \longrightarrow T + F \mid (T) \mid id$$

Removal of Unit Production

$$E \longrightarrow E * T \mid T$$
$$T \longrightarrow E + E \mid (E) \mid id$$
$$T \longrightarrow T + F \mid F$$
$$F \longrightarrow (T) \mid id$$

- We can now replace T with Production-1 such that the next unit production becomes-

$$E \longrightarrow E * T \mid T + F \mid (T) \mid id$$

- Hence, the final productions after removing unit production are:

$$E \longrightarrow E * T \mid T + F \mid (T) \mid id$$
$$T \longrightarrow E + E \mid (E) \mid id$$
$$T \longrightarrow T + F \mid (T) \mid id$$
$$F \longrightarrow (T) \mid id$$

Removal of Useless Productions

- **Useless Symbols:**
- **The non-terminals which do not derive any terminal string from the starting symbol or which do not appear in any sentential form are called useless symbols.**

- **Procedure:**
  - First identify all useless productions and symbols in a CFG and then remove all of them.

Removal of Useless Productions

- **Example:** Remove all useless productions from the following grammar:

$$E \longrightarrow aEb \mid bEa \mid EE \mid ab \mid a \mid B$$
$$A \longrightarrow a$$

Solution:

- In this grammar, we can see that from the starting symbol E, we can reach B, but from B the grammar is not able to generate the string. So, B is a useless symbol.
- Similarly, the non-terminal A is generating the string 'a' but we are not able to reach A from the starting symbol E. Hence, in the given grammar, B is a useless symbol, while A→a is a useless production.
- Therefore, both are simply removed from the grammar, and after removal the grammar becomes:

$$E \longrightarrow aEb \mid bEa \mid EE \mid ab \mid a$$

Normal Forms of CFG

- We know that in a CFG, the RHS of a production may be any string of non-terminals and terminals.
- Besides reducing the grammar it may also be useful for imposing certain restrictions on the form of the remaining productions.
- There are certain standard ways given by different scientists for writing this reduced CFG. These are called normal forms. **So, when the productions in G satisfy certain restrictions then G is said to be in a normal form.**
- There are several normal forms which we can establish for CFG, such as **Chomsky normal form and Greibach normal form.**

Chomsky Normal Form (CNF)

- Noam Chomsky, the creator of the CFG, had imposed the restriction that the RHS of the productions in the CFG can have at most two symbols.
- So, any CFG (free of ε and unit productions) which follows the following restriction on the production rules is said to be in Chomsky normal form (CNF).
- If a grammar is in CNF then the syntax tree or parse tree has at most two descendants at every node: either two internal nodes or a single leaf, as-

$$A \longrightarrow BC$$

$$A \longrightarrow a$$

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

Chomsky Normal Form (CNF)

- Thus. we can say that CNF allows either two non-terminals or only one terminal on the RHS of the production in CFG.
- Any CFG can he converted to CNF by imposing the following restrictions:
- **Rules for converting a CFG into CNF:** The rules that need to be followed for the conversion of CFG into CNF are listed below.
  1) The CFG should not contain null productions or unit productions. So, first remove all null and unit productions.
  2) Remove the terminals from the RHS if it contains more than one terminal. These terminals may be replaced by a new non-terminal.
  3) Restrict the number of non-terminals on the RHS. If any production contains more than two non-terminals, then using the new non-terminal, we may convert the production into CNF.
-

Chomsky Normal Form (CNF)

- **Example: Convert the following CFG into CNF.**

$$A \rightarrow 01XY$$
$$X \rightarrow 1XY \mid \varepsilon$$
$$Y \rightarrow YX0 \mid X \mid \varepsilon$$

- **Solution:**
- First, eliminate all null productions, unit productions, and useless productions.
- **Step 1:** Remove null productions.

  X and Y are nullable non-terminals. Removal of null productions will give us the following grammar.

$$A \longrightarrow 01XY \mid 01X \mid 01Y \mid 01$$
$$X \longrightarrow 1XY \mid 1X \mid 1Y \mid 1$$
$$Y \longrightarrow YX0 \mid X \mid Y0 \mid X0 \mid 0$$

Chomsky Normal Form (CNF)

$A \rightarrow 01XY$

$X \rightarrow 1XY \mid \varepsilon$

$Y \rightarrow YX0 \mid X \mid \varepsilon$

- **Step 2:** Remove unit productions.

  Here, only one unit production Y → X is present. So, We may simply add the production

  $$Y \longrightarrow 1XY \mid 1X \mid 1Y \mid 1$$

  and delete

  $$Y \longrightarrow X$$

- **Step 3:** Apply some new non-terminals, if required, to the reduced grammar and restrict the productions to be in CNF

$$A \longrightarrow 01XY \mid 01X \mid 01Y \mid 01$$

$$X \longrightarrow 1XY \mid 1X \mid 1Y \mid 1$$

$$Y \longrightarrow YX0 \mid 1XY \mid 1X \mid 1Y \mid 1 \mid Y0 \mid X0 \mid 0$$

Let us assume $S \longrightarrow 0$ and $T \longrightarrow 1$ are the productions; now, the above productions will become

$$A \longrightarrow STXY \mid STX \mid STY \mid ST$$

$$X \longrightarrow TXY \mid TX \mid TY \mid 1$$

$$Y \longrightarrow YXS \mid TXY \mid TX \mid TY \mid 1 \mid YS \mid XS \mid 0$$

$$S \longrightarrow 0 \text{ and } T \longrightarrow 1$$

Chomsky Normal Form (CNF)

$$A \longrightarrow STXY \mid STX \mid STY \mid ST$$
$$X \longrightarrow TXY \mid TX \mid TY \mid 1$$
$$Y \longrightarrow YXS \mid TXY \mid TX \mid TY \mid 1 \mid YS \mid XS \mid 0$$
$$S \longrightarrow 0 \text{ and } T \longrightarrow 1$$

- Here, some of the productions remain unchanged because they are already in CNF. Now, we have many productions and their RHS is too long. So, let us assume the following substitutions:

$$R_1 \longrightarrow XY$$
$$R_2 \longrightarrow ST$$
$$R_3 \longrightarrow YX$$

- After substituting these productions we obtain the final CFG which is in CNF as follows:

$$A \longrightarrow R_2R_1 \mid R_2X \mid R_2Y \mid ST$$
$$X \longrightarrow TR_1 \mid TX \mid TY \mid 1$$
$$Y \longrightarrow R_3S \mid TR_1 \mid TX \mid TY \mid 1 \mid YS \mid XS \mid 0$$
$$R_1 \longrightarrow XY$$
$$R_2 \longrightarrow ST$$
$$R_3 \longrightarrow YX$$
$$S \longrightarrow 0 \text{ and } T \longrightarrow 1$$

?

THE END