Parul® University

NAAC A++

Compiler design
Chapter-2: Introduction to syntax analysis

Asst.Prof. Vaibhavi Parikh
Assistant Professor
Department of Computer Science and Engineering

Parul® University

NAAC A++

Content

INDEX

Use of context-free grammars (CFG) in the specification of the syntax of programming languages

- Inherently recursive structures of a programming language are defined by a C
- In a context-free grammar, we have:
  - A finite set of terminals (in our case, this will be the set of tokens)
  - A finite set of non-terminals (syntactic-variables)
  - A finite set of productions rules in the following form
    - $A \rightarrow \alpha$ where A is a non-terminal and
      α is a string of terminals and non-terminals
  - A start symbol (one of the non terminal symbol)

- Example:

  $E \rightarrow E + E \mid E - E \mid E * E \mid E / E \mid -E$

  $E \rightarrow (E)$

  $E \rightarrow id$

Derivations

$E \Rightarrow E+E$

- E+E derives from E
  - we can replace E by E+E
  - to able to do this, we have to have a production rule $E \rightarrow E+E$ in our grammar.

$E \Rightarrow E+E \Rightarrow id+E \Rightarrow id+id$

- A sequence of replacements of non-terminal symbols is called a **derivation** of id+id from E.

- In general a derivation step is

  $\alpha A \beta \Rightarrow \alpha \gamma \beta$ if there is a production rule $A \rightarrow \gamma$ in our grammar

  where α and β are arbitrary strings of terminal and non-terminal symbols

$\alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \alpha_n$ (α_n derives from α_1 or α_1 derives α_n)

$\Rightarrow$: derives in one step

$\Rightarrow$: derives in zero or more steps

$\Rightarrow$: derives in one or more steps

Example

$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E+E) \Rightarrow -(id+E) \Rightarrow -(id+id)$

OR

$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E+E) \Rightarrow -(E+id) \Rightarrow -(id+id)$

- At each derivation step, we can choose any of the non-terminal in the sentential form of G for the replacement.

- If we always choose the left-most non-terminal in each derivation step, this derivation is called as **left-most derivation**.

- If we always choose the right-most non-terminal in each derivation step, this derivation is called as **right-most derivation**.

Left most and Right most Derivation

Left-Most Derivation

$$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E+E) \Rightarrow -(id+E) \Rightarrow -(id+id)$$

Right-Most Derivation

$$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E+E) \Rightarrow -(E+id) \Rightarrow -(id+id)$$

- We will see that the top-down parsers try to find the left-most derivation of the given source program.

- We will see that the bottom-up parsers try to find the right-most derivation of the given source program in the reverse order.

PPT Content Resources Reference Sample:

1. **Book Reference**

   Aho, A. V., Lam, M. S., Sethi, R., & Ullman, J. D. (2006). *Compilers: Principles, Techniques, and Tools* (2nd ed.). Pearson Education.

2. **Journal Article**

   Muchnick, S. S., & Hecht, M. S. (2018). Advances in compiler optimization: Modern approaches for language processing. *Journal of Computer Science and Engineering*, 12(4), 233–245

3. **Website Reference**

   GeeksforGeeks. (2024). *Structure of a compiler*. Retrieved November 7, 2025, from https://www.geeksforgeeks.org/structure-of-compiler/ .

   1. **Conference Presentation**

      Sharma, R., & Patel, D. (2022). *Applications of compiler technology in modern software development*. Paper presented at the International Conference on Advanced Computing and Communication Systems (ICACCS), Chennai, India.

4. **Report**

   IEEE Computer Society. (2021). *Trends in programming language translation and compiler design*.

5. **Sources**

   TutorialsPoint. (2024). *Lexical Analysis in Compiler Design.*

Parul® University | NAAC GRADE A++

https://paruluniversity.ac.in/