

Chapter 1: Regular Expressions in Python

Prof Chauhan Bhavik
MSc IT – Cyber Security | Semester 7

What is a Regular Expression?

- A Regular Expression (Regex) is a sequence of characters that forms a search pattern.
- Used for matching, searching, and manipulating text.
- Powerful for data validation, text parsing, and data cleaning.
- Example: Pattern: `\d{3}-\d{2}-\d{4}` matches 123-45-6789.

Why Use Regular Expressions?

- Automates text searching and manipulation.
- Saves time over manual string handling.
- Used in: Data cleaning, Form validation, Log file parsing, Scraping data.
- Real-World Example: Validate an email address or phone number input.

The re Module in Python

- Python provides built-in support for Regular Expressions via the re module.
- Common functions:
 - `re.match()`
 - `re.search()`
 - `re.findall()`
 - `re.sub()`
 - `re.compile()`

Basic Pattern Examples

- Some commonly used patterns:
 - `\d` → Digit (0-9) → Matches: 7
 - `\w` → Word character (a-z, A-Z, 0-9, _) → Matches: hello123
 - `.` → Any character except newline → Matches: a, 1, @
 - `^abc` → String starts with 'abc' → Matches: abcde
 - `abc$` → String ends with 'abc' → Matches: 123abc