

# Regular Expression Components

## Literal Characters

- **a, b, c, 1, 2, 3, ...:** Any single character that is not a special character is matched literally.

## 2. Metacharacters

These characters have special meanings in regular expressions:

- **.** (**Dot**): Matches any single character except a newline.
- **^** (**Caret**): Matches the start of a line.
- **\$** (**Dollar**): Matches the end of a line.
- **\** (**Backslash**): Escapes a metacharacter to match it literally.
- **[ ]** (**Square Brackets**): Matches any one of the enclosed characters.
- **[ ^ ]** (**Negated Character Class**): Matches any character not enclosed in the brackets.
- **|** (**Pipe**): Acts as a logical OR to match either the expression before or after the |.
- **( )** (**Parentheses**): Groups expressions together, also used for capturing.
- **{ }** (**Curly Braces**): Specifies a number or range of repetitions.

## 3. Quantifiers

Quantifiers specify how many times the preceding element must occur for a match:

- **\***: Matches 0 or more occurrences of the preceding element.
- **+**: Matches 1 or more occurrences of the preceding element.
- **?**: Matches 0 or 1 occurrence of the preceding element.
- **{n}**: Matches exactly n occurrences of the preceding element.
- **{n, }**: Matches n or more occurrences of the preceding element.
- **{n, m}**: Matches between n and m occurrences of the preceding element.

## 4. Anchors

Anchors do not match characters but positions within the text:

- **^**: Matches the start of a line.
- **\$**: Matches the end of a line.
- **\b**: Matches a word boundary (the position between a word and a non-word character).
- **\B**: Matches a non-word boundary.

## 5. Character Classes

Character classes allow matching specific sets of characters:

- **[abc]**: Matches any one of the characters a, b, or c.
- **[a-z]**: Matches any one character from the range a to z.
- **[^abc]**: Matches any character except a, b, or c.
- **[0-9]**: Matches any digit from 0 to 9.

## 6. Predefined Character Classes

These are shorthand notations for common character classes:

- **\d**: Matches any digit, equivalent to **[0-9]**.
- **\D**: Matches any non-digit character, equivalent to **[^0-9]**.
- **\w**: Matches any word character (alphanumeric plus **\_**), equivalent to **[a-zA-Z0-9\_]**.
- **\W**: Matches any non-word character, equivalent to **[^a-zA-Z0-9\_]**.
- **\s**: Matches any whitespace character (space, tab, newline).
- **\S**: Matches any non-whitespace character.

## 7. Groups and Backreferences

- **(abc)**: Captures the group "abc" for use with backreferences.
- **\1, \2, etc.**: Backreferences to the captured groups, where **\1** refers to the first captured group, **\2** to the second, and so on.

## 8. Lookahead and Lookbehind Assertions (Advanced)

These assertions allow you to match a pattern only if it is (or is not) followed or preceded by another pattern:

- **(?=...)**: Positive lookahead; matches if ... follows the current position.
- **(?!...)**: Negative lookahead; matches if ... does not follow the current position.
- **(?<=...)**: Positive lookbehind; matches if ... precedes the current position.
- **(?<!...)**: Negative lookbehind; matches if ... does not precede the current position.

## 9. Non-Capturing Groups

- **(?:...)**: Groups the enclosed elements without capturing them for backreferences.

## 10. Greedy vs. Non-Greedy Matching

- **\*, +, ?**: By default, these are greedy, meaning they match as much as possible.
- **\*?, +?, ??**: Non-greedy versions; they match as little as possible.

## 11. Escape Sequences

Certain characters need to be escaped with a backslash (\) to be treated literally:

- \.: Matches a literal dot.
- \\*: Matches a literal asterisk.
- \[: Matches a literal opening bracket.

## 12. POSIX Character Classes (Used in grep and sed)

- [:alnum:]: Alphanumeric characters.
- [:alpha:]: Alphabetic characters.
- [:digit:]: Digits.
- [:lower:]: Lowercase letters.
- [:upper:]: Uppercase letters.
- [:space:]: Whitespace characters.
- [:punct:]: Punctuation characters.
- [:xdigit:]: Hexadecimal digits.