PW SKILLS

What is regex and Different Character in Regex and importance

Lecture CheckList

1. Introduction.

2. Use cases of regex.

3. The basic structure of regex.

4. Types of regex characters.

Introduction.

Regular expressions which are often referred to as regex or regexp are a powerful tool that helps developers work with strings or text. They are popularly used as a pattern-matching tool which is very much helpful in searching and manipulating specific strings in a text.

Regex is so powerful that we can use regex to match specific characters, words or even the entire string of a text document. They are most commonly used in text editors, and programming languages for tasks such as search, replace, and data validation.

Introduction.

Regex patterns are made up of a combination of characters and special symbols. The most basic regex pattern is a simple string of characters. For example, the pattern "hello" would match the string "hello" in a text document. We will look into many of such patterns in further lectures.

Regex also supports the use of character classes, which we will be looking further, it allows you to match a specific set of characters. For example, the character class "[a-z]" will match any lowercase letter. Additionally, regex provides a number of special character sequences, such as \d to match any digit, \s to match any whitespace, and \w to match any word character.
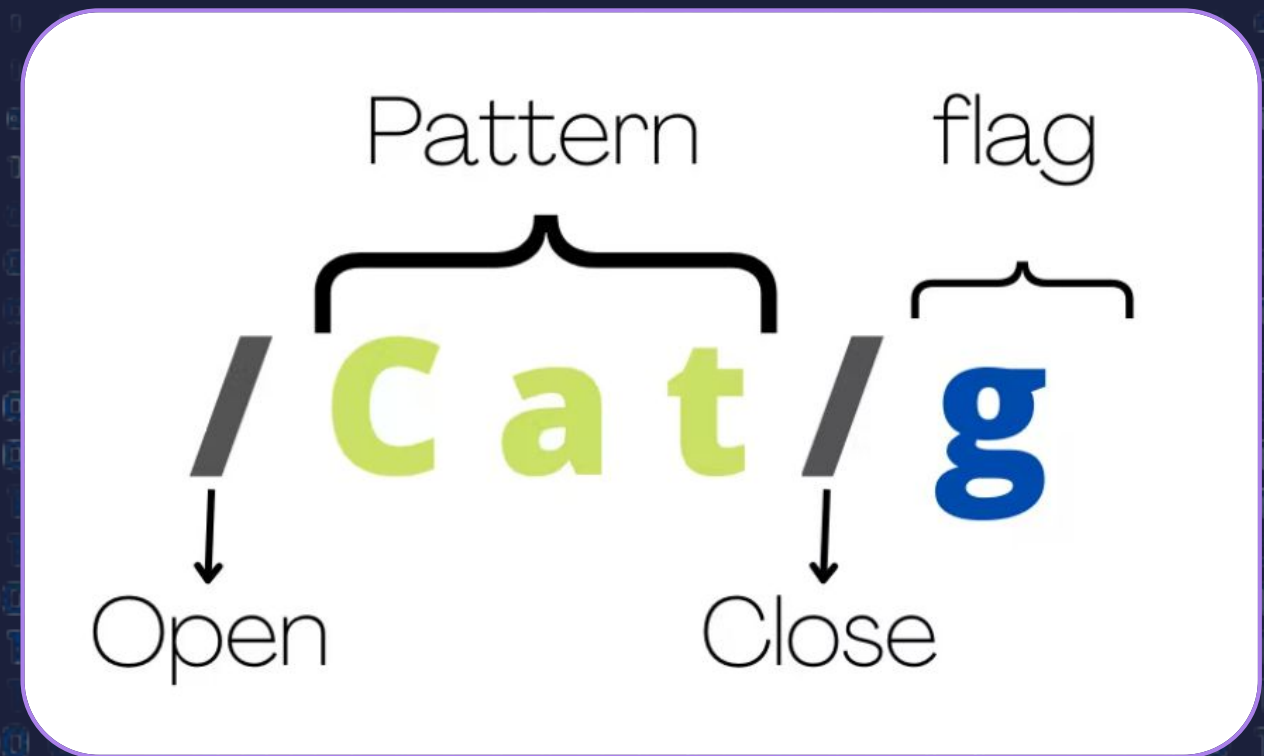
Use Cases of Regex

Some of the most common use cases of a regex are:

- **Search and replace:** You can use regex to search for a specific pattern in a text document and replace it with another pattern. This is often used in text editors and programming languages to make changes to a document.

- **Data validation:** Regex can be used to ensure that a specific string of text conforms to a certain pattern, such as an email address or phone number. This can be used to check user input on a website or in a program to make sure that it is valid.

- **Data Cleaning:** Regex can be used to clean up messy data, such as removing unwanted characters or converting text to a consistent format.

Basic Structure of Regex

The basic structure of regex has the following parts:

1. An opening tag.

2. A closing tag.

3. The pattern.

4. The flag.

Types of regex characters.

The different characters in the regex can be mainly classified into

1. Flags.
2. Character classes.
3. Characters.
4. Anchors.
5. Quantifiers.

Flags

Regex flags are optional options or modifications that can be added to a regex pattern to control its behavior.

Some of the common flags in the regex are

- "g"
- "i"
- "m"
- "s"
- "u"
- "x"

Character Classes

A character class matches any one of a given set of characters. It is used to match the basic elements of a language like a letter, a digit, a space, a symbol, etc.

- "\d"

- "\w"

- "\s"

- "\D"

- "\W"

- "\S"

Characters

- "[abc]"

- "[^abc]"

- "[a-z]"

- "[A-Z]"

- "[0-9]"

- "[set_of_characters]"

- "\"

- "."

- "^"

Anchors

Anchors in regex are special characters that match specific positions within a string, rather than matching specific characters.

Some of the commonly used anchors are:

- "^" (caret)

- "$" (dollar)

Quantifiers

Quantifiers are the repeaters in regex. These are used to specify the number of times a character, character class, or group should be repeated in a match.

- "*" (asterisk)

- "+" (plus)

- "?" (question mark)

- "{}" (curly brackets)