

 Cập nhật tháng 8 năm 2024

## [Bài đọc] Regular Expression - Biểu thức chính quy

Biểu thức chính quy (Regular Expression, thường viết tắt là regex) là một ngôn ngữ nhỏ gọn dùng để tìm kiếm và thao tác trên chuỗi. Trong Java, regex được hỗ trợ mạnh mẽ thông qua các lớp trong gói `java.util.regex`, chủ yếu là `Pattern` và `Matcher`

### 1. Tại sao sử dụng biểu thức chính quy?

- Kiểm tra một chuỗi có tuân theo định dạng mong muốn không (ví dụ: email, số điện thoại, mã bưu điện).
- Tìm kiếm, thay thế các mẫu chuỗi cụ thể trong văn bản.
- Chia tách chuỗi dựa trên mẫu cụ thể.

### 2. Các thành phần cơ bản:

Ký tự đặc biệt trong regex

Ký hiệu	Ý nghĩa	Ví dụ
.	Bất kỳ ký tự nào ngoại trừ ký tự xuống dòng	a.b khớp với acb, a1b
*	Lặp lại 0 hoặc nhiều lần ký tự trước đó	a* khớp với aaa, a, ""
+	Lặp lại 1 hoặc nhiều lần ký tự trước đó	a+ khớp với aaa, a
?	Ký tự trước đó xuất hiện 0 hoặc 1 lần	a? khớp với a hoặc ""
[ ]	Một tập hợp ký tự	[abc] khớp với a, b, c
^	Bắt đầu chuỗi hoặc phủ định trong tập hợp	^a khớp với chuỗi bắt đầu bằng a
\$	Kết thúc chuỗi	a\$ khớp với chuỗi kết thúc bằng a
\d	Một chữ số	\d+ khớp với 123, 456

### 3. Các lớp hỗ trợ Regular Expression trong Java:

#### Lớp Pattern

Pattern đại diện cho một mẫu regex được biên dịch.

- Tạo mẫu regex:

```
Pattern pattern = Pattern.compile(regex: "a*b");
```

#### Lớp Matcher

Matcher là lớp thực hiện thao tác khớp chuỗi dựa trên Pattern.

- Khớp chuỗi:

```
Pattern pattern = Pattern.compile(regex: "a*b");  
Matcher matcher = pattern.matcher(input: "aaab");  
System.out.println(matcher.matches()); // true
```

### 4. Các ví dụ minh họa:

- Kiểm tra định dạng email

```
public static void main(String[] args) new *  
{  
    String email = "example@gmail.com";  
    String regex = "^[\\w-\\.]+@[\\w-]+\\. [a-z]{2,}$";  
    Pattern pattern = Pattern.compile(regex);  
    Matcher matcher = pattern.matcher(email);  
    System.out.println("Is valid email: " + matcher.matches());  
}
```

- Tách chuỗi với regex

```
public static void main(String[] args) new *  
{  
    String text = "apple,banana,grape";  
    String[] fruits = text.split(regex: ",");  
    for (String fruit : fruits)  
    {  
        System.out.println(fruit);  
    }  
}
```

- Tìm kiếm chuỗi con trong văn bản

```
public static void main(String[] args) {
    String text = "I love Java programming";
    Pattern pattern = Pattern.compile(regex: "Java");
    Matcher matcher = pattern.matcher(text);
    if (matcher.find())
    {
        System.out.println("Found at index: " + matcher.start());
    }
}
```

#### 5. Một số chú ý:

- Regex trong Java yêu cầu \ được viết dưới dạng \\ vì ký tự \ là ký tự đặc biệt trong chuỗi Java.
- Việc biên dịch một regex phức tạp nên được kiểm tra kỹ để tránh lỗi hoặc hiệu suất kém.

Link tài nguyên đọc thêm: <https://www.geeksforgeeks.org/regular-expressions-in-java/>

Danh sách các bài học

