**[Bài đọc] Regular Expression**

**LESSON PROGRESS**

**88% Complete**

Biểu thức chính quy là các cú pháp bạn sử dụng để so khớp và thao tác trên các chuỗi. Javascript cung cấp công cụ rất mạnh để làm việc với chuỗi ký tự và các chữ số thông qua việc sử dụng đối tượng RegExp và các ký tự đại diện trong biểu thức chính quy. Ví dụ sau là biểu thức chính quy để so khớp từ JavaScript:

let myRegex = /JavasScript/;

Biểu thức chính quy ở trên so khớp chuỗi JavaScript ở bất kỳ chỗ nào mà nó xuất hiện. Ví dụ nó sẽ khớp trong câu “Đây là sách về JavaScript” và chuỗi “DaylasachveJavaScript” nhưng sẽ không khớp với “Đây là sách về javascript”, vì biểu thức chính quy có phân biệt chữ hoa và chữ thường.

**Cú pháp của biểu thức chính quy**

Các biểu thức chính quy trong JavaScript thường được rút gọn và có cú pháp khó hiểu. Tuy nhiên, bạn không phải lo ngại về chuyện này vì đó thực chất là thế mạnh của biểu thức chính quy. Ví dụ, biểu thức chính quy sau tìm các chữ số và sau đó định dạng lại thành khối điacj chỉ IP ở dạng 192.168.0/24 bằng cách gom nhóm. Đây không phải là một ví dụ phức tạp về biểu thức chính quy.

s/([0-9]+)\.([0-9]+)(\/[0-9]+)/$1\.$2\.0$3/

Biểu thức chính quy trương tự có thể được viết trong JavasScript bằng cách sử dụng hàm replace của đối tượng string, như sau:

let theIP = “192.168.0/28”;

alert(theIP.replace(/([0-9]+)\.([0-9]+)(\/[0-9]+)/, “$1\.$2\.0$3”) );

Cú pháp của biểu thức chính quy bao gồm rất nhiều ký tự có ý nghĩa đặc biệt, bao gồm cả các ký tự đánh dấu so khớp để bắt đầu hoặc kết thúc một chuỗi, ký tự đại diện, ký tự nhóm và một số tự khác.

Bảng sau mô tả ký tự đặc biệt phổ biến trong biểu thức chính quy của JavasScript:

|  |  |
| --- | --- |
| **Ký tự** | **Mô tả** |
| ^ | So khớp với bất kỳ chuỗi nào bắt đầu bằng chuỗi đó. Ví dụ ^n – bất kỳ chuỗi nào bắt đầu bằng ký tự n |
| $ | So khớp với bất kỳ chuỗi nào kết thúc bằng chuỗi đó. Ví dụ $n – bất kỳ chuỗi nào kết thúc bằng ký tự n |
| . | Khớp với bất kỳ ký tự nào |
| \* | Khớp với ký tự trước nó 0 hoặc nhiều lần. Đây là một ký tự đại diện |
| + | Khới với ký tự trước nó 1 hoặc nhiều lần |
| ? | Khớp với ký tự trước nó 0 hoặc một lần |
| () | Nhóm các ký tự khớp vào trong cặp dấu ngoặc đơn để sử dụng sau |
| {n} | Khớp với ký tự trước nó ít nhất n lần |
| {n, m} | Khớp với ký tự trước nó ít nhất n lần nhưng không quá m lần |
| [] | Định nghĩa một tập ký tự để tiến hành so khớp với bất kỳ ký tự nào nằm trong tập này. Ký tự này có thể sử dụng dải như 0-9 để so khớp các chữ số hoặc a-z để so khớp các ký tự |
| [^] | Việc sử dụng dấu ^ trong một tập ký tự sẽ phủ định tập ký tự đó, nghĩa là các ký tự bên trong tập không thể xuất hiện trong chuỗi so khớp |
| \ | Được sử dụng như một ký tự thoát, nghĩa là những gì theo sau dấu gạch chéo được cọi là một ký tự thông thường thay vì có ý nghĩa đặc biệt. Dấu \ cũng có thể được sử dụng để khai báo các bộ ký tự đặc biệt. |

Ngoài ký tự đặc biệt, còn nhiều chuỗi ký tự được dùng để so khớp với các nhóm ký tự hoặc các ký tự không phải chữ số. Một số chuỗi ký tự như vậy được trình bày trong bảng sau:

|  |  |
| --- | --- |
| **Ký tự** | **Chuỗi so khớp** |
| \b | Tìm và so khớp chuỗi nằm sau ký tự \b với những ký tự ở vị trí bắt đầu và kết thúc của một từ. Ví dụ: \bW3 sẽ khớp với chuỗi W3 trong từ W3School. |
| \B | Tìm và so khớp chuỗi nằm sau ký tự \B với những ký tự không nằm ở bị trí bắt đầu và kết thúc của một từ. Ví dụ: chuỗi School trong từ W3School sẽ so khớp với biểu thức chính quy \Bschool |
| \c | Ký tự Control khi sử dụng kết hợp với ký tự khác. Ví dụ \cA là ký tự thoát cho Control-A |
| \d | Chữ số |
| \D | Không phải chữ số |
| \n | Xuống dòng |
| \r | Về đầu dòng |
| \s | Ký tự khoảng trắng (dấu cách hoặc tab) |
| \S | Một ký tự không phải là khoảng trắng |
| \t | Dấu tab |
| \w | Ký tự chữ hoặc số |
| \W | Ký tự không phải là chữ hay số |