

Biểu thức chính qui (Regular Expression)

Biểu thức chính quy (hay còn gọi là regex) bản thân nó là một ngôn ngữ.

Biểu thức chính quy là chuỗi miêu tả một bộ các chuỗi khác, theo những quy tắc, cú pháp nhất định

Dùng chuẩn hóa dữ liệu đầu vào hoặc tìm kiếm theo một định dạng mẫu cho trước.

Biểu thức chính qui (Regular Expression)

Sử dụng nhiều trong Validate để ràng buộc dữ liệu các text box khi nhập và submit trong form:

Cú pháp: /pattern/modifiers trong đó:

pattern: chuỗi Regular Expression

modifiers: thông số cấu hình cho chuỗi với các giá trị:

i: so khớp chữ hoa và thường

g: so khớp toàn chuỗi

m:so khớp cả dữ liệu xuống dòng

Sử dụng hàm test() [true/false]: so khớp biểu thức chính qui

Biểu thức chính qui

Partern chứa 2 loại ký tự:

- Ký tự thường bao gồm các ký tự trong bảng chữ cái: a,b,c,d,... và các chuỗi thông thường được ghép từ các ký tự đó.
- Ký tự đặc biệt:
- ^ : Trả về chuỗi kết quả phù hợp với vị trí đầu của chuỗi gốc.
- \$: Trả về chuỗi kết quả phù hợp với vị trí cuối của chuỗi gốc.
- \ : Ký tự Escape.
- + : Lặp lại 1 hay nhiều lần.

Biểu thức chính qui

Mẫu: $^[a-z0-9_]{3,15}\$$

$^$ Ký hiệu cho biết bắt đầu một dòng

$[a-z0-9_]$ Cho phép tên chứa ký tự a-z, số từ 0 - 9, ký tự -, ký tự _

$\{3,15\}$ Tên dài 3 đến 15 ký tự

$\$$ Điểm kết thúc dòng

Ký tự Meta	Mô tả
.	Biểu diễn bất kỳ ký tự nào ngoài trừ ký tự xuống dòng
[]	Tập hợp ký tự. Phù hợp nếu có bất kỳ ký tự nào trong dấu []
[^]	Tập hợp ký tự phủ định. Phù hợp nếu không có ký tự nào trong []
*	Lặp lại 0 đến nhiều lần.
+	Lặp lại 1 hoặc nhiều lần
?	Tùy chọn có hay không cho mẫu phía trước
{n,m}	Độ dài tối thiểu là n tối đa là m
(xyz)	Biểu diễn một nhóm.
	Biểu diễn thay thế, phép toán or
\	Biểu diễn ký tự đặc biệt [] () { } . * + ? ^ \$ \
^	Điểm bắt đầu của dòng.
\$	Điểm kết thúc của dòng

Tập hợp ký tự []

Dùng [] để chứa tập hợp các ký tự. Có thể dùng dấu - để biểu diễn một dải các ký tự theo vị trí trong bảng chữ cái như a-z, 0-9 ..., biểu thức so sánh sẽ hợp mẫu nếu chứa bất kỳ ký tự nào trong đó (không cần quan tâm thứ tự)

Ví dụ biểu thức [ưƯ]ớc có nghĩa là: Có một chữ ư hoặc Ư, theo sau bởi ó, tiếp theo là c

[ưƯ]ớc => Ước một điều ... mong ước rất đơn sơ kết quả:

Biểu thức chính qui

Expression	Mô tả
[abc]	Tìm các ký tự a, b, c.
[^abc]	Tìm các ký tự không phải a, b, c.
[0-9]	Tìm các ký tự là chữ số từ 0-9
[^0-9]	Tìm các ký tự là không phải chữ số từ 0-9
(x y)	Tìm ký tự x hoặc y
\d	Tìm các ký tự chữ số
\D	Tìm các ký tự không phải chữ số
\s	Tìm các khoảng trắng
\w	Tìm ký tự chữ cái
\W	Tìm ký tự không phải chữ cái
+	Kiểm tra ký tự xuất hiện một hoặc nhiều lần
*	Kiểm tra ký tự xuất hiện không hoặc nhiều lần
?	Kiểm tra ký tự xuất hiện không hoặc một lần
{X}	Kiểm tra ký tự xuất hiện đúng X lần
{X,Y}	Kiểm tra ký tự xuất hiện ít nhất X lần và tối đa Y lần
{X,}	Kiểm tra ký tự xuất hiện ít nhất X lần
^	Kiểm tra ký tự bắt đầu chuỗi
\$	Kiểm tra ký tự kết thúc chuỗi

Biểu thức chính qui

Ví dụ: Biểu thức chính quy cho số CMND:

pattern = `"/^[0-9]{10}$/"`

Biểu thức chính quy cho password từ 6-20 ký tự, con số hoặc các ký tự đặc biệt. pattern = `"/^[A-Za-z0-9(!@#$%^&*()_]{6,20} $/"`

Biểu thức chính quy cho địa chỉ mail:

pattern = `"/^[a-zA-Z0-9]+\@[a-zA-Z]{4,7}+[a-zA-Z]{3}/"`;

Biểu thức chính quy cho số điện thoại di động 10 hoặc 11 số.

pattern = `"/^(01\d{9})|(09\d{8})$/"`;

Biểu thức chính quy cho mã thẻ cào: pattern = `"/^[0-9]{4}-[0-9]{4}-[0-9]{4}-[0-9]{4}$/"`;

Biểu thức chính qui

BIỂU THỨC CHÍNH QUY THƯỜNG DÙNG:

1. Kiểm tra định dạng username:

- Username bao gồm các ký tự chữ cái, chữ số, dấu gạch dưới, dấu chấm độ dài 6-32 ký tự:
- pattern=/[^][A-Za-z0-9_\.]{6,32}\$/[/]

Biểu thức chính qui

BIỂU THỨC CHÍNH QUY THƯỜNG DÙNG:

2. Kiểm tra định dạng password:

Mật khẩu bao gồm các ký chữ cái, chữ số, ký tự đặc biệt, dấu chấm

Bắt đầu bằng ký tự in hoa, độ dài 6-32 ký tự

Pattern=/^([A-Z]){1}([\w_\.!@#\$\$%^&*()]+){5,31}\$/

Biểu thức chính qui

BIỂU THỨC CHÍNH QUY THƯỜNG DÙNG:

3. Kiểm tra định dạng Email

Chứa các ký tự chữ cái, chữ số, dấu chấm, gạch dưới Ký tự @ Nhóm ký tự trước @ có 6-32 ký tự, Nhóm ký tự sau @ là domain một hoặc nhiều cấp

```
pattern=^[A-Za-z0-9_.]{6,32}@([a-zA-Z0-9]{2,12})(.[a-zA-Z]{2,12})+$
```

GHI NHỚ

Biểu thức chính quy là biểu thức tạo ra định dạng mẫu của đối tượng cho trước.

Biểu thức chính quy được xây dựng phức thuộc vào yêu cầu của từng hệ thống

Hàm `pattern.test("textID")` kiểm tra dữ liệu có khớp với biểu thức chính quy hay không

Biểu thức chính qui

Ký hiệu chấm .

Ký hiệu dấu chấm . là một meta đơn giản, nó biểu diễn bất kỳ ký tự nào ngoài trừ ký tự return \r hoặc newline \n. Ví dụ biểu thức .tàn thì có nghĩa là: một ký tự nào đó, tiếp theo đến ký tự t, tiếp theo đến à cuối cùng là n. Ví dụ dùng mẫu đó tìm trong chuỗi.

.tàn => Sự tàn phá môi trường nếu chúng ta không ngăn chặn tàn phá rừng thì rất nguy hiểm

Biểu thức chính qui