

EXPRESSÕES REGULARES – REGEX

Uma expressão regular é um padrão que identifica uma porção de texto. Elas são utilizadas para criar regras com o objetivo de validar um dado, por exemplo, verificar se o que foi digitado é válido de acordo com a natureza do conteúdo do dado. Fazer uma validação significa construir uma regra (expressão regular - regex) e verificar se os dados digitados estão de acordo com a regra.

Uma regra é composta por caracteres especiais que representam o comportamento ou tipo de dado. Alguns exemplos de regras que podemos construir para identificar o dado são: "possui @", "contém apenas números", "é composto por 11 números", "deve conter pelo menos 1 ponto", entre outros. Esses caracteres especiais, que serão usados para construir as regras, são chamados de metacaracteres.

Metacaracteres básicos de uma expressão:

. ? * + ^ \$ | [] { } ()

EXEMPLO DE APLICAÇÃO:

Validação de CPF:

Neste exemplo o propósito é validar se os caracteres digitados são numéricos e se atendem o formato de CPF, ou seja, o CPF é composto apenas por números e, opcionalmente pode ter 2 caracteres especiais: o ponto e o hífen. Isso nos leva a dois padrões que podem ser identificados:

123.456.789-00 (números com pontos e um hífen) ou
12345678900 (apenas números)

Assim, uma das regras para CPF é:

`/^[0-9]{3}\.[0-9]{3}\.[0-9]{3}\-[0-9]{2}$/`

Explicando a regra:

A barra **/** indica o início e fim da regra, o **^** (circunflexo) indica a obrigatoriedade de iniciar os dados apenas com **[0-9]** caracteres numéricos e que sejam os **{3}** primeiros. O ponto tem um significado especial nas expressões regulares, mas no caso do CPF, queremos que o ponto signifique apenas um ponto. Para isso, usamos a barra invertida (****) que é conhecida como um caractere de escape e diz que o caractere que vem na frente deve ser interpretado como um caractere comum e não um metacaractere. Ou seja, o ponto a seguir é parte do CPF. Neste caso o ponto é opcional. Isso é indicado pela interrogação seguida do ponto: **?**, o mesmo serve para o hífen no final da expressão que também é opcional: **-?**.

No final da regra, **[0-9]{2}** indica que se espera encontrar apenas números e que eles devem ser exatamente duas repetições.

O problema desta regra é que se digitado um dos formatos a seguir, será considerado correto, pois tanto ponto (.) como hífen (-), são opcionais.
123456789-00, 123.456789-00, 123456.78900

Assim, para atender aos dois formatos será necessário modificar e ampliar a regra:

/^([0-9]{3}.[0-9]{3}.[0-9]{3}-[0-9]{2}) | ([0-9]{11})\$/

A barra vertical | (pipe) representa o operador lógico **OU**. Na expressão ou regra acima define que um CPF deve conter **números, pontos e hífen** ou apenas **números**:

123.456.789-00 ou **12345678900**

Tabela de Metacaracteres

Metacaracter	Descrição
/	Indica início e fim de expressão regular.
^	Indica uma pesquisa no início do bloco/regra.
\$	Indica uma pesquisa no fim do bloco/regra.
[]	Definir uma classe/lista de caracteres.
[^]	Utilizado para negar uma classe/lista de caracteres.
()	Defini um bloco/clausula de caracteres.
{x}	Repetir exatamente x de vezes.
{x,y}	Repetir no mínimo x ou no máximo y de vezes.
?	Declara um ou mais caracteres como opcional.
+	Repete pelo menos uma vez.
	Representa o conector logico (ou).
.	Indica qualquer caractere.
*	Indica nenhuma ou várias ocorrências de um ou vários caracteres numa regra.
\d	o mesmo que [0-9]
\D	o mesmo que [^0-9]
\s	espaço, quebra de linha, tabs etc.; o mesmo que [\t\n\r\f\v]
\S	o mesmo que [^ \t\n\r\f\v]
\w	o mesmo que [a-zA-Z0-9_]
\W	o mesmo que [^a-zA-Z0-9_]