

# Expressões Regulares

João Marcelo Uchôa de Alencar

Universidade Federal do Ceará - Quixadá

24 de agosto de 2023

# O que é uma Expressão Regular?

- ▶ Uma **Expressão Regular** (ER) é uma construção que utiliza um conjunto finito de operadores para representar cadeias de caracteres em um alfabeto que possuem um formato regular.
- ▶ **Expressões Regulares** são escritas em uma linguagem formal que pode ser interpretada por um processador de expressões.

# Para quem servem Expressões Regulares?

- ▶ Busca e substituição de textos em editores e linguagens de programação.
- ▶ Validação de formatação de texto.
- ▶ Realce de sintaxe.
- ▶ Filtragem de informação.

# Operações Básicas

## Alternância

Uma barra vertical (|) separa alternativas. A expressão 'casa|Casa' pode aceitar tanto a cadeia 'casa' quanto 'Casa'.

## Agrupamento

Parênteses ( ) são usados para definir escopo e procedência. A expressão '(c|C)asa' é equivalente à anterior.

## Quantificação

Um quantificador após um *token* ou agrupamento especifica a quantidade de vezes que o elemento precedente pode ocorrer:

- ▶ ? indica zero ou uma ocorrência.
- ▶ \* indica zero ou mais ocorrências.
- ▶ + indica uma ou mais ocorrências.

Chamamos ?, \*, + de metacaracteres.

# Metacaracteres

Os metacaracteres se classificam em:

- ▶ Âncoras
- ▶ Representantes
- ▶ Quantificadores

ER	Função
<code>^</code>	Pesquisar texto no início das linhas
<code>\$</code>	Pesquisar texto no fim das linhas
<code>\b</code>	Pesquisar no início ou fim das palavras
<code>\B</code>	Negação de <code>\b</code>

## grep - Pesquisa arquivos por conteúdo

```
$ grep [-opções] [expressão] [arquivo1] [arquivo2] ...
```

<b>-E</b>	<b>Ativa suporte estendido à expressões regulares</b>
<b>-a</b>	Força a tratar o arquivo como texto
<b>-c</b>	Apenas informar quantas linhas contém a expressão
<b>-i</b>	Não diferenciar maiúsculas e minúsculas
<b>-l</b>	Apenas informar qual dos arquivos contém a expressão
<b>-v</b>	Busca reversa
<b>-n</b>	Exibe o número da linha

# Representantes

ER	Função	
.	Ponto	Qualquer caractere uma vez
[]	Lista	Qualquer dos caracteres
[^]	Lista negada	Nenhum dos caracteres da lista



# Quantificadores

ER	Função	
?	Opcional	Torna a entidade anterior opcional
*	Asterisco	Zero ou mais ocorrência
+	Mais	Uma ou mais ocorrências
{ }	Chaves	Quantidade exata

# Classes POSIX

<code>[ :alnum: ]</code>	Alfanuméricos	<code>[ A-Za-z0-9 ]</code>
<code>[ :alpha: ]</code>	Alfabéticos	<code>[ A-Za-z ]</code>
<code>[ :blank: ]</code>	Espaços	<code>[ \t ]</code>
<code>[ :cntrl: ]</code>	Controle	<code>[ \x00-\x1F\x7F ]</code>
<code>[ :digit: ]</code>	Dígitos	<code>[ 0-9 ]</code>
<code>[ :graph: ]</code>	Visíveis	<code>[ \x21-\x7E ]</code>
<code>[ :lower: ]</code>	Minúsculas	<code>[ a-z ]</code>
<code>[ :print: ]</code>	Visíveis e Espaços	<code>[ \x21-\x7E ]</code>
<code>[ :punct: ]</code>	Pontuação	<code>[ .... ]</code>
<code>[ :space: ]</code>	Espaços, mais nova linhas	<code>[ \t \r \n \v \f ]</code>
<code>[ :upper: ]</code>	Maiúsculas	<code>[ A-Z ]</code>
<code>[ :xdigit: ]</code>	Hexadecimais	<code>[ A-Fa-f0-9 ]</code>

# Grupos

Os parênteses permitem criar grupos de caracteres que são avaliados como um só.

`(vice-)?(governa|sena|verea|)dora?`

Quantas palavras casam com essa expressão?

O grupo no parênteses pode ser referenciado no resto da expressão.

`([A-Za-z]+\1`

Toda e qualquer sequência de caracteres repetida.

## Exemplo - Parte I

Considere o arquivo *emailsordenados.txt*. Vamos criar um arquivo chamado *parte\_01.sh* e colocar nele, um por linha e na ordem abaixo, os seguintes comandos:

1. Um comando *grep* que encontre todos os alunos cujo primeiro nome começa com A.
2. Um comando *grep* que encontre todos os alunos que tem algum nome começando com A.
3. Um comando *grep* que encontre todos os alunos cujo e-mails terminam com '.br'.
4. Um comando *grep* que encontre todos os alunos cujo e-mails contém números.

## Exemplo - Parte II

Quais seria as expressões regulares que representam as seguintes cadeias de caracteres?

1. A palavra "revista" no singular e no plural.
2. A palavra "letra" em qualquer combinação de letras maiúsculas ou minúsculas.
3. Os números inteiros.
4. Um número IP.