

Expressões Regulares

João Marcelo Uchôa de Alencar

Universidade Federal do Ceará - Quixadá

22 de Agosto de 2019

O que é uma Expressão Regular?

- ▶ Uma **Expressão Regular** (ER) é uma construção que utiliza pequenas ferramentas para obter uma determinada sequência de caracteres de um texto.
- ▶ **Expressões Regulares** são escritas em uma linguagem formal que pode ser interpretada por um processador de expressões.

Para quem servem Expressões Regulares?

- ▶ Procura e substituição de textos em editores e linguagens de programação.
- ▶ Validação de formatos 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

- ▶ 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] ...
```

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

Atividades - Parte I

Crie na sua pasta o diretório *atividades/atividade03*. Copie para ele o arquivo */home/compartilhado/emailsordenados.txt*. Crie um arquivo chamado *parte_01.txt* e coloque 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'*.

Atividades - Parte II

Copie o arquivo `/home/compartilhado/expressoesregulares.txt` para a pasta `atividades/atividade03`. Crie um arquivo `parte_02.txt` com os comandos `grep`, um por linha, para as seguintes buscas:

1. Descubra quais linhas tem um espaço em branco antes da pontuação final.
2. Descubra quais linhas tem um sinal de pontuação seguido diretamente de uma letra, sem espaço entre ela e o sinal.
3. Descubra quais linhas não tem uma frase começando com letra maiúscula após a pontuação.
4. Descubra quais linhas não terminam com pontuação.

Atividades - Parte III

Crie o arquivo *atividades/atividade03/parte_03.txt* e coloque, uma por linha, 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.

Atividades - Parte IV

Considerando o arquivo *expressoesregulares.txt*, coloque em um arquivo *parte_04.txt* com um comando *grep* para a seguinte busca:

- ▶ Descubra quais linhas contém palavras com mais que 8 caracteres.