

## 103.7: Expressões regulares, grep e sed

Rafael Obelheiro

28/08/2024

## Expressões regulares (1)

- Usadas para diversas tarefas de processamento de texto em ferramentas de linha de comando, editores de texto e linguagens de programação
- O padrão POSIX divide em duas classes
  - ERs básicas → obsoletas, herdadas de ed
    - ★ para manter compatibilidade com ferramentas antigas
  - ERs estendidas (modernas, usadas por egrep)
- A realidade é mais complicada do que isso
  - existem pequenas diferenças entre as sintaxes aceitas pelas várias ferramentas/linguagens que suportam ER
  - inicialmente, grep usava ERB, e egrep foi introduzido para dar suporte a ERE
    - ★ hoje, as ERBs suportadas pelo GNU grep têm elementos de EREs
  - sempre é bom consultar uma tabela de compatibilidade

## Expressões regulares (2)

### Special characters in regular expressions (common ones)

Symbol	What it matches or does
.	Matches any character
[chars]	Matches any character from a given set
[^chars]	Matches any character not in a given set
^	Matches the beginning of a line
\$	Matches the end of a line
\w	Matches any "word" character (same as [A-Za-z0-9_])
\s	Matches any whitespace character (same as [\f\t\n\r]) <sup>a</sup>
\d	Matches any digit (same as [0-9])
	Matches either the element to its left or the one to its right
(expr)	Limits scope, groups elements, allows matches to be captured
?	Allows zero or one match of the preceding element
*	Allows zero, one, or many matches of the preceding element
+	Allows one or more matches of the preceding element
{ n }	Matches exactly <i>n</i> instances of the preceding element
{ min, }	Matches at least <i>min</i> instances (note the comma)
{ min,max }	Matches any number of instances from <i>min</i> to <i>max</i>

a. That is, a space, a form feed, a tab, a newline, or a return

Figure 1

## grep

- Principal comando de busca de ERs, pode processar arquivos ou ser usado como filtro
- Sintaxe: `grep [opções] [regex] [arq1] [arq2] ... [arqN]`
  - se não for especificado nenhum arquivo, processa stdin

## grep: opções mais comuns

opção	significado
-------	-------------

-E	seleciona ERE (mesmo que egrep)
-F	casas strings fixas (mesmo que fgrep)
-f	obtem ERs de um arquivo, 1 por linha (OR)
-c	conta linhas
-i	ignora maiúsculas/minúsculas
-l	lista apenas os nomes de arquivos que casam com a ER
-n	mostra números de linha
-v	mostra linhas que <b>não</b> casam com o padrão
-H	mostra nome do arquivo (útil com find)

## sed: stream editor

- Filtro que realiza operações de manipulação de texto
  - adicionar, remover, alterar texto
- Faz uso extensivo de ERs
  - sed -E permite o uso de EREs
- O uso mais comum na prática é

```
$ cat arq | sed 's/foo/bar/g' # foo -> bar
```