

EXPRESSÕES REGULARES

ADMINISTRAÇÃO DE SISTEMAS
2020/2021

ROLANDO MARTINS
(ADAPTADOS DE PEDRO BRANDÃO)

Referências dos slides

- O conteúdo destes slides é baseado no livro da disciplina: “Unix and Linux System Administration Handbook (4ªEd)” por Evi Nemeth, Garth Snyder, Trent R. Hein e Ben Whaley, Prentice Hall, ISBN: 0-13-148005-7
- As imagens usadas têm a atribuição aos autores ou são de uso livre.

RE – Onde são usadas?

- Para encontrar padrões complexos, mas abrangentes
 - `grep`, `vi`, `awk`, `sed`, `perl`. etc.
- `perl` tem expressões regulares avançadas
 - No `grep` pode-se fazer `-E` para as ter (ou usar `egrep`)
 - No `sed` pode-se fazer `-r` para as ter

Matching

- Verdadeiro se se encontrar o padrão definido no texto dado
 - Padrão: Administração de Sistemas
 - ✓ Texto: A disciplina de Administração de Sistemas é top.
 - ✗ Texto: A disciplina de administração de sistemas é mais cenas.
 - Sensibilidade às maiúsculas e minúsculas

MATCHING

5

[dcc]

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])
\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 n instances of the preceding element
{ min, }	Matches at least min instances (note the comma)
{ min,max }	Matches any number of instances from min to max

EXEMPLOS

- Padrão: Administração (de)?
(Sistemas|Redes)

✓ Texto: Administração de Sistemas é o máximo

✓ Texto: A disciplina de Administração de Redes é top

✗ Texto: A disciplina de Administração Redes é cool

- P: up[0-9]{9}@[^.]+.up.pt

○ up201307429@fc.up.pt

○ up2017429@fc.up.pt

○ up201307429@fe.up.pt

○ UP201307429@fc.up.pt

○ up201307429@dcc.fc.up.pt

○ up201307429@fc.upept

○ up201307429@fc.uc.pt

○ up201307429@fc.up.pttt

Exemplo do Livro

- P: M[ou]'?am+[ae]r ([AEae]l[-])?[GKQ]h?[aeu]+([dtz][dhz]?) +af[iy]
- Variantes da escrita de Moammar Gadhafi
 - Muammar al-Kaddafi (BBC)
 - Moammar Gadhafi (Associated Press)
 - Muammar al-Qadhafi (Al-Jazeera)
 - Mu'ammar Al-Qadhafi (U.S. Department of State)

Captura

- Dados entre parênteses são capturados
 - ~~Em perl ficam em \$1, \$2, etc.~~
 - Em grep ou sed ficam \1, \2, etc.
- ~~Exemplo em perl de retirar espaços entre “-” do código postal (no ficheiro fich_codP)~~

~~—~~

```
perl -i'.bak' -pe 's/(\d{4})\s*-\s*(\d{3})/$1-$2/' fich_codP
```




Gulodice vs. preguiça



- Ao usar quantificador + ou * a expressão tenta apanhar o máximo. (guloso)
- Usando +? ou *? Tenta apenas até suceder (preguiçoso).
- grep e sed não fazem a preguiça...
 - `sed -E 's/(.+?):.* /\1/' /etc/passwd`
 - `grep -E '.*+?:' /etc/passwd`
 - (ver colorido)
- ~~Em perl sim:~~
 - ~~`perl -ne '/(.+?):/; print "$1\n";' /etc/passwd`~~

Resumo

- Expressões regulares são muito poderosas
 - Podem também ser muito complexas
- Padrão e texto
- Corresponder, quantificar, agrupar
- Captura
- Preguiça e gulodice

QUESTÕES/ COMENTÁRIOS