



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale  
2007-2013



Platformă de e-learning și curriculum e-content  
pentru învățământul superior tehnic

Utilizarea Sistemelor de Operare

30. Filtre de text

## Filtre de text

- All is text. (Unix)
- Comenzi care “filtrează” text
  - `filtru < input_file > output_file`
  - `command1 | filtru | command2`
- În Windows PowerShell se lucrează cu obiecte
- Operații
  - selectare linii
  - selectare coloane (după separator)
  - prelucrare linii (modificare elemente)
  - contabilizare

## cat, tac, rev, nl

- cat
  - afișează intrarea standard la ieșirea standard
  - Ce efect are rularea comenzii cat fără nici un argument?
    - CTRL+D
  - **cat -n**
- tac
  - afișare inversată (prima linie este afișată ultima)
- rev
  - afișare inversată a fiecărei linii (primul caracter este afișat ultimul)
- nl – number lines

## head, tail

- **head -10, head, head -n 10**
  - primele 10 linii
- **head -n -20**
  - toate liniile mai puțin ultimele 20
- **tail -10, tail, tail -n -10**
  - ultimele 10 linii
- **tail -n +21**
  - toate liniile începând de la linia 21 (adică mai puțin primele 20)

# cut

- Selectare coloane de text
- Separatorul implicit este TAB
- `cut -f 1 < file`
- `cut -f 1-3 < file`
- `cut -f 1,4 -d ':' < /etc/passwd`

## tr

- Transliterate
- Transformare litere în alte litere
  - `tr 'a' 'b' < file`
  - `tr 'a-z' 'A-Z' < file`
- Ștergere
  - `tr -d ' ' < file`
- Squeeze
  - `ifconfig | tr -s ' ' | cut -d ' ' -f 2`
- `tr -d -c 'A-Za-z' < /dev/urandom | head -c 8`

## sort, uniq

- Sortarea liniilor
- Implicit alfabetică
- `sort -n file` – sortare numerică
- `sort -r file` – reverse sort
- `sort -u file, sort file | uniq`
- `sort -k 3 file` – sortează în funcție de coloana 3

## WC

- Contorizarea elementelor
- Cu argument fișier afișează **și numele fișierului**
- **wc -l** – numărul de linii
- **wc -c** – numărul de caractere
- **wc -w** – numărul de cuvinte



# grep

- Extrage linii ce conțin un anumit pattern
- Folosește expresii regulate
- `grep 'pattern' file`
- `grep -r 'pattern' /path/to/folder/` - căutare recursivă
- `grep -v 'pattern' file` – invert match
- `grep -i 'pattern' file` – ignore case
- `grep --color=auto`
  - recomandăm folosirea unui alias

## Expresii regulate. grep

```
razvan@anaconda:/tmp/for$ grep '^ana' /etc/passwd
ana:x:1118:1026:Ana Gainaru:/home/students/ana:/bin/bash
razvan@anaconda:/tmp/for$ grep 'ana' /etc/passwd
list:x:38:38:Mailing List Manager:/var/list:/bin/sh
roxana:x:1004:1004:Roxana Savulescu,,,:/home/users/roxana:/bin/bash
roxanam:x:1030:1026:Roxana Murarus:/home/students/roxanam:/bin/bash
constantin:x:1079:1028:Constantin Blanariu:/home/students/constantin:/bin/bash
[...]
```

```
razvan@anaconda:/tmp/for$ grep -r '^#include <stdio\.h>$' /usr/include
/usr/include/argp.h:#include <stdio.h>
/usr/include/mntent.h:#include <stdio.h>
/usr/include/printf.h:#include <stdio.h>
```



## Expresii regulate (2)

- $x^*$  - caracterul  $x$  de oricâte ori (posibil niciodată)
- $x^+$  - caracterul  $x$  de oricâte ori (cel puțin o dată)
- $x?$  – caracterul  $x$  cel mult o dată
- $.$  (punct) – orice caracter
- $^$  - început de rând
- $\$$  - sfârșit de rând
- $[a-z]$  – set de caractere
- $[^a-z]$  – set complementar de caractere
- $a|b$  – caracterul  $a$  sau  $b$
- **Atenție** – la grep unele trebuie escapate ( $\backslash+$ ,  $\backslash|$ )

## Expresii regulate – Exerciții

- Expresie pentru o dată în formatul YYYY-MM-DD
- Expresie regulată pentru adrese MAC (0C:A3:11:CF:D1)
- Expresie regulată pentru nume valide Windows
  - Nu încep cu spațiu sau punct
  - Nu conțin \, /, \*, ?, :, ", <, >, |
- Expresie regulată pentru un URL HTTP
  - `http://www.google.com/a%b?d=45`
  - considerați că URL-ul se încheie la apariția unui spațiu sau virgulă

## find și xargs

```
razvan@anaconda:/tmp$ find /etc -type f -name '*.conf'  
/etc/mkinitrd/mkinitrd.conf  
/etc/host.conf  
/etc/nsswitch.conf  
[...]
```

```
razvan@anaconda:/tmp$ find /etc/ -type f -exec stat -c '%s' {} \;  
469  
380  
1029  
[...]
```

```
razvan@anaconda:/tmp$ find /etc/ -type f | xargs stat -c '%s'  
469  
380  
1029  
[...]
```