

Introduction to the Bash shell on Linux and Mac OS

Filtering and Processing Text

Reindert-Jan Ekker

@rjekker

<http://www.linkedin.com/in/rjekker>



Overview

- Editing text files
- Sorting
- Searching
- Filtering
- Replacing text
- Finding files
- Column-based files

Editors

- **Nano**

- Tiny, simple, fast
- Pre-installed on Mac and most Linux distro's

- **Vi (vim)**

- Very full-featured, but still light-weight
- Pre-installed on Mac and Linux
- Press "i" for insert mode
- Press "Esc" for command mode
- :w to save file
- :q to exit
- :q! to exit without saving
- Start "vimtutor" to learn more

Editors (2)

■ Emacs

- Very full-featured
- Pre-installed on Mac OS and some Linux distro's
- F10 for menu
- Get out with ^X ^C
- To get help: start it without arguments

Sorting

- **Command: sort**
 - Sorts file alphabetically
- **-r reverses sort**
- **-n sorts numerically**
- **-k sorts by field**
 - `sort -k 2`
 - space-separated fields by default
- **Filter out repeated lines: uniq**
 - `sort attendance | uniq`
 - `-c` counts lines

Head and Tail

■ Head

- Show first 10 lines of input by default
- -n gives number of lines
- `head -n 1`

■ Tail

- Show last 10 lines of input by default
- -n gives number of lines
- `tail -n 1`
- `tail -f` follows newly appended data

Grep

- **Grep searches for a text in a file**
 - `grep string file`
- **Or multiple files**
 - `grep string *`
- **Important options:**
 - `-i` makes search case-insensitive
 - `-c` counts occurrences
 - `-l` shows line number of occurrence
 - `-v` inverts the search (filter)
- **Can use regular expressions**

Find

- **Very powerful tool for searching files**
- **find dir**
 - Lists all files in dir
- **find dir -name f**
 - Lists all files named f in dir
 - Match expression:
 - `find dir -name '*txt'`
- **Find has many powerful options**
- **For more info: <http://goo.gl/Rid5>**

Advanced tools

- **sed**

- Stream editor
- Transform text
- Replace words
- Most common use: to replace “old” with “new”:
- `sed 's/old/new/g'`

- **awk**

- Complete programming language
- Very useful for column-oriented files

- **Perl**

- **Python**

Column-based data

- **sort -k to sort on a specific column**
 - use -t to specify delimiter
- **cut to select a column from the input**
 - cut -f 2 shows only second column
 - -d for delimiter
- **paste**
 - puts lines of input files next to each other in the output
- **join**
 - joins input files based on matching keys

Summary

- **Background jobs**

- &
- ^Z, bg, fg

- **Inspect processes**

- jobs
- ps
- top

- **End processes**

- kill
- xkill
- pkill

Summary

- **Editing text files**
 - Nano, vi, emacs
- **Sorting**
 - sort, uniq
- **Head and tail**
- **Searching and filtering**
 - grep
- **Replacing text**
 - tr, sed
- **Finding files**
 - find
- **Advanced tools**
 - awk, sed, perl, python
- **Column-based files**
 - cut, paste, join
- **Input redirection with <**
- **Multiple commands on one line**
 - ;