Unix
Command-Line Processing

L435/L555

Dept. of Linguistics, Indiana University
Fall 2014
What is Unix?

Unix is an operating system, like DOS or Windows

- developed in 1969 by Bell Labs
- works well for single computers as well as for servers
- underlying operating system for Macs

Linux: open source version of Unix (mostly for PCs)
Directory Structure

directory = folder

▶ directory structure: tree structure

Important directories:

▶ home directory: your private directory (/Users/guest)
▶ root node: /
▶ important directories on a Mac:
  ▶ /Applications
  ▶ /Users
  ▶ /Volumes
  ▶ /Volumes/Data/
Graphical Representation

```
/       Applications      Users      Volumes
     |     md7       guest     ...     Data
     |                |           ar en de multilingual ...
```

Unix

Command-Line Processing

Directories

Files

Useful Commands

Permissions

tar

More Commands

Working Remotely

Emacs

4 / 22
Useful Commands for Directories

- list files in current directory:
  `ls`

- list files with more information:
  `ls -l`

- change directories:
  `cd <to>`

  two possibilities to specify `<to>`:
  - from root:
    `cd /Users/md7`
  - from where we are:
    `cd ../md7`
    `..` means 'go up'

- go back to home directory:
  `cd`

- show the directory where you are:
  `pwd`
Useful Shortcuts

**TAB completion**
If you start typing a command or filename, then press TAB, the shell will complete the word for you – as far as possible.

**Command history**
The shell keeps a history of your commands. To scroll through them, simply press the up arrow key.
Looking at Files

- display contents of file in terminal:
  `cat <filename>`

- display page by page:
  `less <filename>`
  - next page: space bar
  - quit: q
  - go to beginning of file: g
  - go to end of file: G
  - search forward: `/<expression>` + hit return
  - search backward: `?<expression>` + hit return
  - in search: next found occurrence: n

- Likewise, there is the `more <filename>` command
Looking at Files

More Commands

- display beginning of file (normally 10):
  head <filename>
- display X beginning lines of file:
  head -X <filename>
- display end of file (normally 10):
  tail <filename>
  - tail -X <filename>: display X ending lines
  - tail +X <filename>: display all lines starting at X
- count number of lines, words, characters in a file:
  wc <filename>
More Useful Commands

- list files that end in ’.txt’:
  `ls *.txt`

- make directory:
  `mkdir <path/><dirname>`
  subdirectory: `mkdir results`
  in your home directory: `mkdir ~/results`

- copy a file from one location to another:
  `cp <filename> <path/>`

- copy a file to the directory ’results’ in your home directory:
  `cp dates.txt ~/results`
More Useful Commands (2)

- sort your file:
  `sort <filename>`

- sort (numbers) in numeric order:
  `sort -n <filename>`

- find a word in a file and display all the lines in which it occurred:
  `grep <word> <fromfile>`

- find a word in a file and display all the lines in which it occurred and save results into a file in your home directory:
  `grep <word> <fromfile> > ~/<tofile>`

- example:
  `grep linguistic mycorp.txt > ~/res.txt`
What other options are there for sorting?

The `man` command lets you see documentation on whatever Unix command is in question

- Find information:
  ```
  man <command>
  man sort | less
  ```
Tangent: grep & egrep

grep is commonly used, but doesn’t offer a full RE syntax

▶ For that, use egrep (extended grep)

Bonus: what does grep stand for?
More Useful Commands (3)

- delete lines (adjacent) with identical content:
  `uniq <filename>`

- delete lines (adjacent) with identical content, but list how many were there:
  `uniq -c <filename>`

- display differences between two files
  `diff <filename1> <filename2>`
Permission System

- `ls -l`

```
-rw-r--r-- 1 root admin 168724 Nov 9 2003 Jokes.pdf
drwxr-xr-x 15 root admin 510 Aug 31 2006 ar
drwxr-xr-x  3 root admin 102 Aug  4 2004 cs
drwxr-xr-x  4 root admin 136 Mar  2 2007 de
drwxr-xr-x  35 root admin 1190 Feb 18 15:23 en
```

- change permissions:
  
  `chmod <who><+/-><what> <file/dir>`
  
  `<who>`: user = u; group= g; others = o; all = a;
  `<what>`: read = r; write = w; execute = x;

- example: give user write permissions for file 'hello.txt':
  `chmod u+w hello.txt`

- example: deny others write and execute permissions for directory 'test':
  `chmod o-wx test`
Packing and Unpacking Files

- **pack a file:**
  - `gzip <file>`
    - results in `<file>.gz`

- **unpack file:**
  - `gunzip <file>.gz`

- ’pack’ more than one file:
  - `tar cvf <resultfile>.tar <file or directory>`
    - unpack:
      - `tar xvf <file>.tar`

- **tar and gzip:**
  - `tar cvzf <resultfile>.tgz <file|directory>`
    - unpack:
      - `tar xzvf <file>.tgz`
More on Copy and Move

- copy a file from one location to another:
  \texttt{cp <path/filename> <path/filename>}

- move a file from one location to another:
  \texttt{mv <filename> <path/>}

- rename a file:
  \texttt{mv <oldfilename> <newfilename>}

- remove a file:
  \texttt{rm <filename>}

- remove a directory including contents:
  \texttt{rm -r <dir>}
Important Shortcuts in Paths

- .. (go) up one directory
- . here
- ~ home directory
Translate

- exchange characters for others:
  \[ \text{tr} \ 'aeiou' \ 'X' \ < \ <\text{filename}> \]
  \[ \text{tr} \ 'aeiou' \ 'AEIOU' \ < \ <\text{filename}> \]

- collapse characters (squeeze):
  \[ \text{tr} \ -s \ 'J' \ < \ <\text{filename}> \]

- exchange complement of characters for others:
  \[ \text{tr} \ -c \ 'a-z' \ '
  \<\text{filename}> \]
  \n is a linebreak
Putting Commands Together

- `< use following filename as input file
- `> use following filename as output file
- `| use output of preceding command as input for following commands

example:
```
cat <infile> | tr 'A-Z' 'a-z' | sort > <outfile>
```
Remote Login

▶ login:
ssh <username>@<machinename>

e.g.  ssh guest@nlp.indiana.edu

▶ logout:
exit
Remote Copy

- to copy a local file to (your home directory) on another machine:
  \texttt{scp <filename> <machine>}: \\
e.g. \texttt{scp vm.pos nlp.indiana.edu}:

- to copy a file from another machine to the current directory:
  \texttt{scp <machine>:\langle path\rangle/\langle filename\rangle} . \\
e.g. \texttt{scp nlp.indiana.edu:/Volumes/Data/en/penntreebankv3/readme.all} . \\
(in one line!)
Emacs

- emacs is a fairly basic text editor that can be run in a window or in the shell
- to start emacs:
  emacs <filename>
- to quit:
  Ctrl-x Ctrl-c
- save:
  Ctrl-x Ctrl-s
- search:
  Ctrl-s

Other shell editors include vim and nano