

LING 300 - Topics in Linguistics:
Introduction to Programming and Text Processing for Linguists

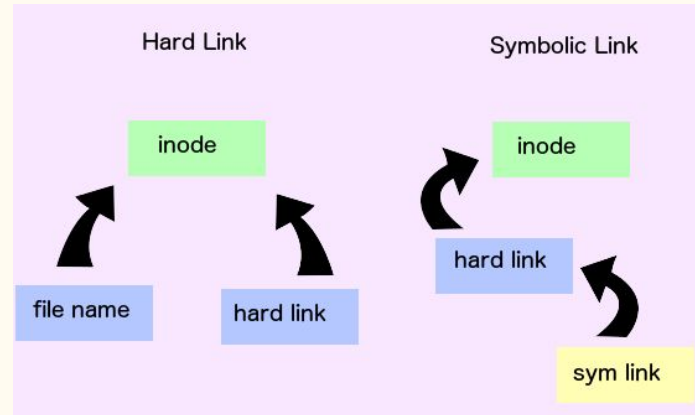
Week 2

—

Command-line Text Processing

Notes from Assignment 1

- Careful with assignment naming so I know what to look at (just edit the existing file)
- Don't remove the ">>>" or "YOUR ANSWER HERE" stuff
- Softlinks vs. Hardlinks



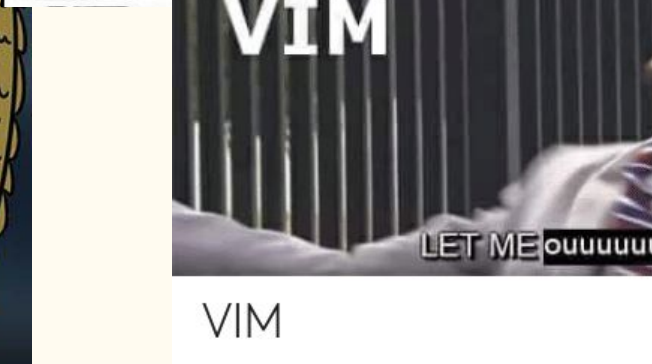
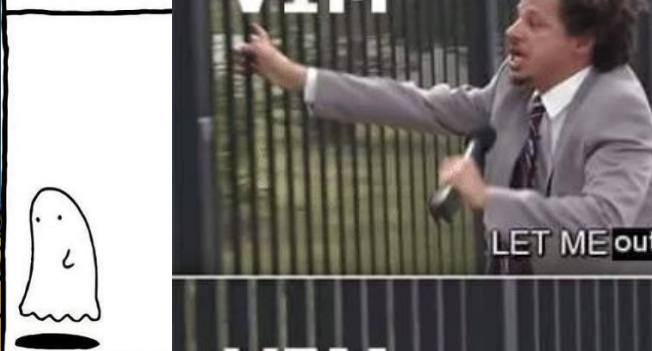
Notes from Assignment 1

- Is the assignment really ~270 lines?
 - Seems maybe more?
- How does `wc` count lines?
 - Returns number of `<newline>` characters, aka `\n`
 - Try `-S` and `-N` flags on `less`

Notes from Assignment 1

- Dotfiles (`.bashrc`, `.nanorc`)
 - Saves us from typing e.g. `umask 002` every time we open a terminal
- Tab completion! Our very good friend.
 - ``cd<tab>`` vs. ``cd <tab>``

Notes from Assignment 1



Important Concepts This Week

- Unix operates line-by-line
- You're creating a data pipeline, and ordering matters
- Imagine starting with a big piece of clay (all the lines) and bit by bit carving away and transforming it
- One liners! These techniques are very powerful for getting a quick, high-level picture of what you're working with



Unix Philosophy

- One program does one thing

Write programs that do one thing and do it well.

Write programs to work together.

Write programs to handle text streams, because that is a universal interface.

-Doug McIlroy (inventor of pipes)

Abstraction

(we can trust programs and functions to do their thing)

We don't need to worry about how something is sorted, we just trust that it will be properly done!

Decomposition

(big problems can be solved by breaking them into small steps)

For each problem, think about how to decompose it into small things we can do

Notes on Assignment 2

- The assignment file itself is an executable shell script
(a type of runnable program)
- The `.sh` ending is a suggestion of this
(your browser may take the suggestion and auto-download)

Notes on Assignment 2 (cont.)

- **Write the command that produces the answer,**
not the answer itself

- The scripting section (section 7) has been tricky in the past,
read carefully and slowly there! *(demo)*