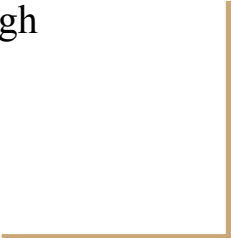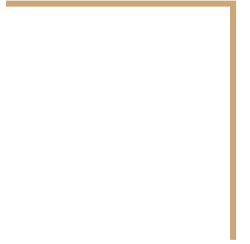# The command line

NEH Institute: Advanced Digital Editing
University of Pittsburgh
July 13, 2022

Review

Review exercises (on the command line)

01. Go to your pr-app repository
    a. Hint: `cd`
02. Add a new file called "foo.txt"
    a. Hint: `touch`
03. Verify that "foo.txt" is there
    a. Hint: `ls`
04. "foo.txt" is an empty file. Add the following text into it: "testing, testing"
    a. Hint: `echo [text] > [filename]`
05. Check that "foo.txt" is no longer empty
    a. Hint: `cat`
06. Add some more stuff. Append "1, 2, 3" to "foo.txt," then repeat step 05
    a. Hint: `echo [text] >> [filename]`
07. Our testing is complete! Get rid of our "foo.txt" file (cautiously!).
    a. Hint: `rm -i`, then `Y`

Getting further acquainted

Expanding our vocabulary

- Patterns in file names
  - `ls [character(s)]*` → Find all files that **begin** with a specific character (or characters)
  - `ls *[character(s)]` → Find all files that **end** in a specific character (or characters)
  - `ls *[character(s)]*` → Find all files that **contain** a specific character (or characters)
  - Example: `ls ghost*` → Returns all files that **begin** with the word (character string) "ghost"
- Counting
  - `wc [file]` → Find the number of lines, words, and bytes in a file (or files)
  - `wc -l [file]` → Find the number of **lines** in a file (or files)
  - `wc -w [file]` → Find the number of **words** in a file (or files)
  - `wc -c [file]` → Find the number of **bytes** in a file (or files)
  - Example: `wc -l ghost*` → Returns the number of lines in each file that begins with "ghost"

Practice Exercises

01. Go to the directory where your local copy of the pr-app repository is located, then go into "data," then "hoax_xml" (the corpus of documents)
    a. Hint: `cd, cd, cd` (until you get to "hoax_xml")
02. Verify that you are in "hoax_xml"
    a. Hint: `pwd`
03. Return the documents that were published in 1838
    a. Hint: `ls *[characters]*`
04. Return the documents that were published in the 1830s
    a. Hint: slightly modify your solution to Exercise 03
05. Find the number of lines in each document published in 1838
    a. Hint: `wc -l  *[characters]*`

Combining lists and counts (and more!)

- `ls` returns a list of files/directories
- `wc -l` counts the number of lines in its input
- How do we combine them? → The pipe operator |
- Pipe (|) → Uses the output of a preceding (left of the pipe) command as the input of a subsequent (right of the pipe) command
- Example: `ls | wc -l`
  - Translation: Count the number of lines in the returned list of directories
  - Answers the question: How many files are in a directory? Subdirectories in a directory?
- Example: `ls *183* | wc -l` → Returns the number of documents published in the 1830s
- Example: `wc -l *183* | sort` → Returns the number of lines per document published in the 1830s, in ascending order

Practice Exercises

01. Return the number of documents in the corpus whose titles contain the word "ghost"
    a. Hint: `ls *[characters]* | wc -w`
02. Find the longest document published between 1800 and 1810
    a. Hint: `wc -l *[characters]* | sort`
03. Find the shortest document published in the 1830s
    a. Hint: `wc -l *[characters]* | sort -r`

"Reading" documents

- `less [filename]` → Returns a scrollable version of a file's contents
- `less -N [filename]` → Does the above, with the addition of line numbers
- `head [filename]` → Returns the **first** few lines of a file's contents
- `tail [filename]` → Returns the **last** few lines of a file's contents
- `grep [pattern] [filename]` → Searches for the pattern within a file (or multiple files)
- `grep -n [pattern] [filename]` → Does the above, with the addition of line numbers
- Example: `grep -n -E "(C|c)onstable" *183*` → Returns the occurrences (with line numbers) of the word "Constable" or "constable" in documents published in the 1830s

Practice Exercises

01. What do you see when you enter `less -N aghost_age_1832.xml`?
02. What do you see when you enter `head aghost_age_1832.xml`?
03. When you repeat Exercise 02 with `tail` instead of `head`?
04. Return the lines that contain the word "constable" in the documents published in the 1830s
    a. Hint: `grep -n -E [pattern] *[characters]*`
05. How would you return the count of the number of lines grep returns?
    a. Hint: pipe `(|)`, `wc -l`

Further resources

- *Learn Enough Command Line to Be Dangerous*
    - https://www.learnenough.com/command-line-tutorial
- tldr pages
    - https://tldr.sh/