

# Digital Humanities at the command line

Jonathan Blaney and Ryan Heuser

What is it and why are you  
making us learn it?

“You should be able to  
control the entire computer  
from the command line.”

IT person at my old job



```
0:25@Pictures:~$ convert melon-head.jpeg
```

```
18:26@Pictures:~$ convert melon-head.jpeg -resize 50% s  
mall-melon-head.jpeg █
```

- 'glue' skill (and sometimes prerequisite knowledge)
- extensibility
- scalability
- speed
- availability
- preference
- DH cheat code (although it's not actually cheating)

Let's run some commands

```
ls -al --color . *rc
```



```
cd Doc
```

```
[press tab to complete]
```

```
touch test1.txt  
test2.txt test3.txt
```

```
mkdir temp
```

```
cp *.txt temp/
```

```
mv file1.txt temp/  
file99.txt
```

```
rm temp/* .txt
```

```
head -n1000  
GUTINDEX.ALL | tail -n5
```

```
grep -i "Quixote\|Quijote"  
GUTINDEX.ALL -A3 | grep -B3 "Spanish"
```



Using `gutenberg.tsv`:

Find all Spanish editions of Don Quixote

Find all Spanish language books

Find all books by Shakespeare in French or German

(don't worry about noise in the results)

Git Bash start  
Mac open  
Linux xdg-open

```
grep -i "Shakespeare" gutenbergr.tsv |  
grep "French\|German" > shakespeare.tsv  
    && open shakespeare.tsv
```

```
grep -i "German\|  
French" gutenbergl.tsv  
| grep "Shakespeare"  
| awk '{print $NF}'
```

```
awk '{print $NF}'
```

```
cut -f2 gutenbergl.tsv |  
sort | uniq -c | sort -nr |  
head -n20 | nl
```

```
cut -f2 gutenbergl.tsv | grep -i  
    "jane austen" | sed 's/,  
January-March 1817// ' | uniq -c
```

Tips and tricks/best practice



Where are the programs?

```
echo $PATH
```

```
python3 --version
```

```
which rm
```

```
which -a python3
```

Best practice: do not edit your  
path manually!

Use your history  
ctrl-r

or

```
history 0 | grep  
history | grep
```

Interrupting, quitting,  
suspending...

`ctrl-c` stop the process

`ctrl-d` stop input in a REPL

`q` quit from paged content

## Some keyboard shortcuts

`ctrl-a` beginning of line

`ctrl-e` end of line

`ctrl-k` delete everything forward

`ctrl-w` delete one word backwards

`ctrl-u` delete line

There are many more, and newer,  
commands that you can install



We need to talk about Vim

To get into insert mode: i

To save and exit Vim

escape :w :q

```
curl -L0 https://www.gutenberg.org/  
ebooks/2000.epub3.images --output  
don-quixote.epub
```

```
#!/bin/bash
curl -L0 https://www.gutenberg.org/  
ebooks/2000.epub3.images --output don-  
quixote.epub
```

```
#!/bin/bash
# download an ePub file from Gutenberg
curl -L0 https://www.gutenberg.org/ebooks/  
2000.epub3.images --output don-quixote.epub
```

```
#!/bin/bash
# download an ePub file from Gutenberg
set -euo pipefail
curl -L0 https://www.gutenberg.org/ebooks/  
2000.epub3.images --output don-quixote.epub
```

```
#!/bin/bash
# download an ePub file from Gutenberg
set -euo pipefail
curl -L0 https://www.gutenberg.org/ebooks/
$1.epub3.images --output $2.epub
```

```
#!/bin/bash
# download an ePub file from Gutenberg
# $1 = the Gutenberg ID
# $2 = the name of the downloaded epub, without termination
set -euo pipefail
curl -L0 https://www.gutenberg.org/ebooks/$1.epub3.images
--output $2.epub
```



# When not to use the command line

- Across platforms (just for you but a different OS)
- In teams
- Anything long term or complicated, or that needs to be secure

Return the lines from *Paradise Lost* where the first word is one of the 10 most common first words in a line in the poem beginning with C

- Work in groups and leave no one behind!
- No scripts, Python, ChatGPT etc. But you can do a web search.
- No need for regex or commands we haven't used today (but you might need flags we haven't used)
- Assume that all first words are proper case, eg *Cheddar* not *CHEDDAR*
- You don't need to do it in one line
- You don't need to finish, this is about practice and teamwork
- If you get really stuck ask Jonathan and Ryan for hints

# If you finished that pretty quickly...

- Do the same exercise with the most frequent last words beginning with any letter
- Make a script where the user can enter nth word in the line and get the most frequent ones
- Make a script where the user can enter the nth word and any poetry text file
- Using only the command line, can you make the GUTINDEX.ALL files in a TSV with only author, title, language and ID per line? How far can you get and what are the blockers?