

# Faillogs

## FindingData

May 26 2018

jordenlee

26 May 2018

in hist3814o

## Exercise 1: The Dream case

The search for

Province = Provinzen

Province = Italic Regions

Italic region = Roma

Search string 1 = figlina

resulted in 25 hits --> displayed chart distribution by province, country, monument, inscription etc...

## CWGC

Last Name: Lee

Filter on WWII and China(including Hong Kong)

URL: [https://www.cwgc.org/find/find-war-dead/results?](https://www.cwgc.org/find/find-war-dead/results?lastName=Lee&tab=wardead&fq_warliteral=2&fq_country=China_+(including+Hong+Kong))

[lastName=Lee&tab=wardead&fq\\_warliteral=2&fq\\_country=China\\_+\(including+Hong+Kong\)](https://www.cwgc.org/find/find-war-dead/results?lastName=Lee&tab=wardead&fq_warliteral=2&fq_country=China_+(including+Hong+Kong))

Step4:

1.opened the nano text using the following in the command line then pasted the above at the top

2.exited and saved in the FileManager... starting to think I goofed this saving function in the first week. because i don't remember if i did, and my faillogs are not indicating what happened the previous week. partly because i kind of forgot about the faillogs last week untill i was further into the modules.

3.“Lodge a copy” seems like the codes i was using didnt work so i manually did uploaded them through github

## Ian Milligan's wget tutorial at the Programming Historian

26 May 2018

in hist3814o

wget -r --no-parent -w 2 --limit-rate=20k <http://activehistory.ca/papers/>

ok... seems like this step has just performed what the above was mentioning (download rate, where to get the pushes from etc...) Seems to be downloading a lot of papers or stuff at a much slower rate just as described.

Faillogs

FindingData

jordenlee

26 May 2018

in hist3814o

get <http://activehistory.ca/papers/>

Some stuff happened. new lines were automatically written. right below are similar messages found on my computer (saving to: 'index.html.1...')

Faillogs

FindingData

jordenlee

26 May 2018

in hist3814o

wget-activehistory

“Command not found”

Faillogs

FindingData

jordenlee

26 May 2018

in hist3814o

mkdir wget-activehistory

This is the code we should type into the command line after we have established whether the installation was correct or not.

Faillogs

FindingData

## Kellen Kurschinski wget tutorial

jordenlee@1b8704a4d5eb:~\$ nano urls.txt

jordenlee@1b8704a4d5eb:~\$ ls

dhbox-work-today.md Git-push-Repo pandoc-1.19.2.1-1-amd64.deb

today'scommands.docx war-diary

dhbox-work-today.mdYes Historic R [urls.py](#) wget-activehistory

first-repo index.html Spoon-Knife urls.txt wget-activehistorymkdir

jordenlee@1b8704a4d5eb:~\$ history

1 history

2 mkdir war-diary

3 cd war-diary

4 pwd

5 cd ~

6 nano [urls.py](#)

7 python [urls.py](#)

8 ls

9 urls.txt

10 nano urls.txt

11 ls

12 history

## Exercise 3 TEI

26 May 2018

in hist3814o

in our supporting materials

took a bit to get going but reading through and following the directions closely should yield an oddly satisfying and rewarding result of seeing the page in color.

Faillogs

jordenlee

26 May 2018

in hist3814o

blanktemplate.txt

This can be found in the Zip file you downloaded in the previous step...

Faillogs

FindingData

TEI

## Exercise 4 APIs

jordenlee

hist3814o

Just now

Make sure to make a file noting what you've done, commands you've made, etc, and upload it in your GitHub repository.

Once you type history in to the command, you will actually have to look for it. Initially i thought it would be in the first level in my repo but it turns out it can be found in the "m2e4" directory/branch.

Faillogs FindingData APIs

## Exercise 5: Mining Twitter

jordenlee@1b8704a4d5eb:~\$ nano out.csv

jordenlee@1b8704a4d5eb:~\$ history

1 search.json

```
2 twarc search hist3814o > search.json
3 jtwarc search hist3814o > search.json
4 twarc configure
5 pip install https://github.com/DocNow/twarc/archive/v1.2.0.tar.gz
6 twarc configure
7 twarc search hist3814o > search.json
8 twarc stream
9 sudo npm install json2csv --save -g
10 ls
11 nano out.csv
12 history
13 history > dhbox-work-today.md
14 ls
15 twarc stream
16 search.json
17 json2csv -i input.json -f carModel,price,color -p
18 json2csv -i search.json -o out.csv
19 ls
20 nano out.csv
21 history
```

## Exercise 6: Tesseract

### (annotations)

Compare these two files. Did the OCR conversion get progressively worse? How do they compare to each other, to the first attempt at conversion, and then to the originals?

As far as i can tell aside from spacing and zoom, they seem to be the same

Faillogs

FindingData

Tesseract

jordenlee

27 May 2018

in hist3814o

```
sudo apt-get install libcurl4-gnutls-dev
```

Is anyone else having troubles here? I don't think this code is fully installing everything it needs to. which I believe is partly the reason why the stuff in R does install correctly either

Faillogs

FindingData

Tesseract

jordenlee

27 May 2018

in hist3814o

```
install.packages
```

none of these commands work for me. i get this message

```
ERROR: dependency 'curl' is not available for package 'magick'
```

```
removing '/home/jordenlee/R/x86_64-pc-linux-gnu-library/3.4/magick'
```

```
Warning in install.packages : installation of package 'magick' had non-zero exit status
```

And when i proceed through the rest of the codes the same thing happens

for each.

As a result, the txt file doesn't get written then I cannot compare the two files in step 15.

Faillogs

FindingData

Tesseract

jordenlee

27 May 2018

in hist3814o

sudo apt-get install libpoppler-cpp-dev

Got a bunch of failed to fetches but just continued on to the next step

Faillogs

FindingData

Tesseract

jordenlee

26 May 2018

in hist3814o

Let's convert the first file to TIFF with ImageMagick's convert command

Were we supposed to save a photo somewhere? because I get a "no images defined" message

This results in step 6 also not being able to proceed as it cannot open the e001518087.tiff file

Faillogs



FindingData

Tesseract

jordenlee

26 May 2018

in hist3814o

`sudo apt-get install tesseract-ocr`

Following the 3 step previously. This command will read and unpack something... Asks to y/n if want to proceed. I clicked y and it prepared and unpacked some stuff

Faillogs

FindingData

Tesseract