

CSCI 4180: Big Data Analytics CSCI/DASC 6010: Big Data Analytics and  
Management Spring 2023  
Homework 1 – Linux power tools

Farnoosh Koleini

B01342500

- 1) (5 pts) List the Linux command (wget or curl), complete with all options and arguments, used to download all the files from this website.

Answer:

```
wget -r -np http://www-cs-students.stanford.edu/~blynn/gitmagic/
```

- 2) (15 pts) List the one line command used to count how many files were downloaded, along with its output (the number of files). Hint: you will have to use multiple commands and pass the output from one command as input to the next command.

Answer:

```
find . -type f | wc -l
```

```
185
```

- 3) (30 pts) List the one line command used to count how many files of each type (extension) were downloaded, along with its output (count and file extension pairs).

Answer:

```
find . -type f -name '*..*' -not -name '.*' | sed -Ee 's,.*./\.[^/]+$,\1,' | sort | uniq -ci | sort -n
```

```
1 epub
```

```
13 css
```

```
13 pdf
```

```
156 html
```

- 4) (50 pts) List the one line command used to find the file with most numbers of lines, along with its output (lines count, and file name and path).

Answer:

```
find . -type f -exec wc -l {} \; | sort -n | tail -1
```

```
2059 ./intl/ko/book.pdf
```