LING 1340/2340 – Data Science for Linguists

26 March 2020

Jevon Heath's Whirlwind Tour of bash

0) Preliminaries

* Go to http://www.gutenberg.org/ebooks/11 and download the Plain Text UTF-8 version, saved to your Desktop. Rename it 'alice.txt'.
* Go to http://norvig.com/ngrams/ and download 'enable1.txt', saved to your Desktop.

1) Navigation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| pwd | pwd | | | |
| cd | cd ~ | | pwd | |
| <TAB> | cd [original folder] | | cd .; pwd | |
| ls (-lhA) | ls | ls -a | ls -A | ls -l |
| (up-arrow and down-arrow) |  | | | |
| <Ctrl>+C | asdf;lkajsdf;lkajsdf;l | |  | |
| <Ctrl>+L |  | |  | |

[How do you handle spaces in file names?]

2) Customization

|  |  |
| --- | --- |
| alias | alias calc='/c/windows/system32/calc.exe' |
|  | alias |
|  | alias ls='ls -lhAF --color=tty' |
|  | alias grep='grep --color' |
| .bash\_profile (login) | cd ~; nano .bash\_profile |
| .bashrc (non-login) | nano ~/.bashrc |

3) Searching

|  |  |
| --- | --- |
| whereis | whereis pip |
| which | which pip |
| echo | echo $PATH |
| locate | locate alice.txt |
| find | find ~ -name 'alice.txt' |

4) Moving files

|  |  |  |
| --- | --- | --- |
| touch | touch dummy.txt | |
| cp | cp dummy.txt dummier.txt | |
| mv | mv ~/Desktop/alice.txt ./alice.txt | |
| rm | rm dummier.txt | |
| mkdir | mkdir folder | |
| rmdir | rmdir folder | |
| rm -v | mkdir folder; touch folder/dummy.txt | |
|  | rmdir folder | rm -rv folder |

5) Examining text files

|  |  |
| --- | --- |
| wc | wc alice.txt |
|  | wc -l alice.txt |
| head (-n) | head -8 alice.txt |
| tail (-n) | tail -6 alice.txt |
| cat | cat alice.txt |
| more or less | less enable1.txt (Space, b, q to navigate) |

6) Piping

|  |  |
| --- | --- |
| > | echo "I am a cat" > cat.txt |
| | | cat alice.txt | wc > alicecount.txt |
| >> | echo "He is too" >> cat.txt; cat cat.txt |

7) Regular expressions

|  |  |
| --- | --- |
| grep (-in) | grep "^o.\*o$" enable1.txt |
| dos2unix | dos2unix enable1.txt |
| grep -P | grep -P '[aeiou]{5,}' enable1.txt |
| grep -v | grep 'q' enable1.txt | grep -v 'u' |
| Plus piping: | grep '^un.\*able$' enable1.txt > able.txt |
|  | grep '^in.\*able$' enable1.txt >> able.txt |
| grep -l | grep -l 'Alice' \*.txt |

8) Text manipulation

|  |  |
| --- | --- |
| sort (-r) | sort cat.txt |
| cut | cp Licensed-Data-Sets/ETS\_Corpus\_of\_Non-Native\_Written\_English/data/text/index.csv ./ETSindex.csv |
|  | cut -d, -f3 ETSindex.csv > ETSlangs.txt |
| uniq (-c) | sort ETSlangs | uniq -c |
|  | head -3375 alice.txt | tail -3345 | perl -npe 's/\s+/\n/g' | sort | grep '\S' | uniq -c | sort -nr | head -30 |

9) Other useful commands – check them out!

|  |  |  |  |
| --- | --- | --- | --- |
| man | do and done | diff | file |
| ps | ( & ) | info | date |
| tac | history | ssh | ln |
| time | awk and sed | bind | join |
| apropos | lsof | fold | type |