Assignment 1

December 10, 2014

Your boss has given you several files that need to be modified and summarized. Your elite shell scripting skills will save the day! Develop a shell script that can achieve the goals listed below, and email Jon the final shell script. **Be sure to read and follow all steps**, and include lots of useful comments in the script.[1] You can work with one other person in the class on this.

The files are named `file_00.txt.gz`, `file_01.txt.gz`, etc.

1. Use `wget` to download all files from `http://languagemodel.org/classes/uds/shell_and_python_basics/hw1_files/`

2. Use `gunzip` to uncompress each file

3. Use `cat` to compile all uncompressed files into a single file[2]

4. Use `iconv` to convert the text from the old-school ISO-8859-1 to modern UTF-8 encoding

5. Convert all uppercase letters to lowercase

6. Use `grep` to find out how many lines contain "Twitter". There is a command-line argument to grep that will simply tell you how many lines match, rather than print the lines that match. Also find out how many lines contain "Google" and "Facebook".

7. Find out what the top 10 most fequent words are that *start* with "euro". Find out the top 10 words *ending* in "platz".

---

[1]Start a line with the `#` symbol, followed by your useful comments.
[2]You can use `zcat` instead if you want.