

ADS1 exercises – Repository and working with text files

1. Create a github account (www.github.com). Create a repository and upload your code(s) from time to time.
2. Attached the file `clean_up.py`. Download it into the folder you will be using to store your code. `import clean_up` under a short name of your choice. Yes, it is that simple to create a module in python. The module contains the function `clean`. Use the help function to find out its purpose and how to call it.
 - (a) Read in the data science text book `big_data.txt` and perform a word frequency analysis. How does the number of different words compare with Jane Austen's 7279?
 - (b) Count the number of words in the text. This can be done in one line using the `len()` function.
 - (c) Count the number of characters. Again you can use the `len()` function to derive the length of the string.
 - (d) Analysis of the frequency of letters is a method to crack simple replacement tools ($a \rightarrow x$, $b \rightarrow p$, $c \rightarrow d$, ...). Carry out a frequency analysis of letters. The `list()` function applied to a string creates a list of characters (print one example).
 - (e) Determine the length of the longest word in the text. Print out the number of characters and the word.