

# NLP HW 3 readme

Aasish Dhungel

May 2021

The programming portion of the homework was done in jupyter notebook because the input file was too large and some editors such as VSCode, and sublime text were getting, "memory error"  
The .py file is also included in the zip folder, however jupyter notebook(ipynb) is recommended.

**NOTE: news.crawl.bz2 is NOT included in the zip folder**

## Jupyter notebook

- 1) To run the file, 'cd' into the directory where the homework file as well as 'news.crawl.bz2' is saved from the terminal.  
*example: 'cd homework\NLP\hw3'*
- 2) Then in the terminal type, 'jupyter notebook'.
- 3) Jupyter notebook should open in your default browser window.
- 4) In the jupyter notebook window, select the file named, 'NLP\_HW3\_Aasish\_Dhungel.ipynb'.
- 5) Run each individual cell after opening the file. The total compile time of the file was approx. 23 minutes.

**NOTE: There were some issues using a text editor and some 'MEMORY ERROR' because the crawl file was too large  
It is recommended to run this file in jupyter notebook as that was more stable for me.**

## Python file

- 1) To run the '.py' file, 'cd' into the directory where the homework file as well as 'news.crawl.bz2' is saved from the terminal  
*example: 'cd homework\NLP\hw3'*
- 2) In the terminal type, 'python nlp\_hw3\_aasish\_dhungel.py'

3) Wait for the file to compile, the compile time for the file was approx. 25 minutes and possible memory error.