

## Week 3 NLP Assignment

### Spelling Corrector

Submitted by:

Surabhi Trivedi

Preet Khowaja

"In `gen_corrupted.py`, I have written a function `spell_check`, which takes an input string and outputs a string with the spellings corrected. In some cases, this function works well. For example, "Pjark" gets corrected to "Park". My spell-checker also knows not to try and match numbers to words in the dictionary. However, it does fail in certain instances. The biggest is its inability to recognize proper nouns and names. It corrects "Dashwood" to the dictionary word with which it has the smallest edit distance. This can be fixed by including a regex capturing group for proper nouns and names. It will not compare these words to the words in our dictionary list.

"Furthermore, if the word in Austen's text is 'Aed' which should be corrected to 'And' but our spell-correcter corrects this word to 'and' instead because in our dictionary 'and' is ranked above 'And'. In some cases when a word is contextually misspelt the spell-correcter will not correct it because it is still a valid word. For example, if 'his' is spelt as 'hi' the spell-checker will not know that this is an incorrect spelling.

"Austen uses language that is dated according to the dictionary we are using and that is another challenge the spell-checker faces when it tries to correct the word but is not able to match the word well.