

Exercise

- 1) Download, Clean, Normalize, Tokenize and POS Tag the following text data:

<https://www.gutenberg.org/files/11/11.txt>

- 2) Compute:
 - a) 10 most frequent word unigrams
 - b) 10 most frequent word bigrams
 - c) 10 most frequent word trigrams
 - d) 10 most frequent POS unigrams
 - e) 10 most frequent POS bigrams
 - f) 10 most frequent POS trigrams