## Homework 1

1. Install Python (if you don’t have it already) and install NLTK.   
   Installed.
2. Follow the instructions in chapter 1 of Bird-Klein for implementing a “lexical diversity” scoring routine.

Completed.

1. Go to <http://www.gutenberg.org/ebooks/bookshelf/215> and select texts of different grade levels (e.g., fourth reader, fifth reader et al)

Report the lexical diversity score of each. Explain whether the result was surprising.

The books I tried all have similar diversity score which is expected as the text are big in books. For really simple kids book, the diversity score is less.

Text

Description automatically generated with medium confidence

1. Also compare the vocabulary size of the same three texts. Explain whether the result was surprising.
   1. Vocabulary size of Animal book 85
   2. Vocabulary size of Pride and Prejudice book 94
   3. Vocabulary size of Wizard of Oz book 90
   4. Vocabulary size of Grade 1 Stories book 86

The more complex book, larger is vocab size.

1. Write a paragraph arguing whether vocabulary size and lexical diversity in combination could be a better measure of text difficulty (or reading level) than either measure is by itself.

Each, vocabulary size and lexical diversity measures layers of difficulties that are expressed in different format, and not each on its own captures the text difficulty. So, combining both has a better measure of text difficulty.