# CIS 4930 NLP // HW #2 // Spring 2018

**Date Assigned:** January 19, 2018
**Date Due:** January 25, 2018

## Submission Format

You will submit a soft copy of your solution using e-Learning ( http://elearning.ufl.edu ) by the end of the day ( 23:59 / 11:59 PM ) on the assigned date ( January 25 ). Save your solution as a **pdf** file and name file **hw2** ( hw2.pdf ).

## Assignment

At the top of every solution file you submit this semester include: your name, section number, the assignment number, and the date due. Complete these exercises. In your answers, you may find it useful to write some code, run a test, and report the result of the test. In addition to reporting test results, analyze your results and assert why you have drawn the conclusions given in your answers.

## Exercises

- **1.5** Compare the lexical diversity scores for humor and romance fiction in 1.1 (http://www.nltk.org/book/ch01.html#tab-brown-types). Which genre is more lexically diverse?

- **1.6**: Produce a dispersion plot of the four main protagonists in *Sense and Sensibility*: Elinor, Marianne, Edward, and Willoughby. What can you observe about the different roles played by the males and females in this novel? Can you identify the couples?

- **1.19**: What is the difference between the following two lines? Which one will give a larger value? Will this be the case for other text?

```
>>> sorted(set(w.lower() for w in text1))
>>> sorted(w.lower() for w in set(text1))
```

These exercises are from Chapter 1 ( http://www.nltk.org/book/ch01.html ). Recall, the numbering system used in the second (online/target copy) edition of the textbook leaves out some prefixing. So the heading **8 Exercises**, the final section of Chapter 1, is Section 1.8. For clarity, the problem statement for each exercise is repeated here.