CIS 4930 NLP // EC #1 // Spring 2018

Date Assigned: March 14, 2018
Date Due: March 19, 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 (March 19). Submit one file, **ec1.py**.

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 the following exercise to receive extra credit on Exam #1.

Exercises

1. [20 pts] Create the function *chi_square*. The function will implement Pearson's Chi Square test. Your function will receive three string values, the two words of a potential collocation and the corpus – for example ('new', 'companies', brown) where brown is the brown corpus (*from nltk.corpus import brown*) – the function will perform the full summation calculation (not the short cut) to deduce the test result. Your solution will print all statistics related to your calculation, c(w1), c(w2), c(w1 && !w2), and so on as well as the Chi Square value and whether or not you have found a collocation (the null hypothesis rejected or accepted). Recall Chi Square uses a baseline of 3.841 for 0.05% accuracy.

Output Structure:

```
C(w1):           #
C(w2):           #
C(w1w2):         #
C(w1 && !w2):    #
C(!w1 && w2):    #
C(!w1 && !w2):   #
Total Words:     #

0.05% Baseline: 3.841
X^2:             #

Whether or not we have a collocation based upon this test.
```