### Assignment on Chapter 3(b) in “*Natural Language Processing with Python --- Analyzing Text with the Natural Language Toolkit” by* Steven Bird, Ewan Klein, and Edward Loper

You will find Chapter 3 posted at:

<http://www.nltk.org/book/ch03.html>

Section 3.12 lists exercises. You will be asked to complete some of the exercises. On the Canvas, a file named

nltkChapter3b\_template.py

is posted. You are requested to use this template in formulating your code solutions.

Let *SimpleText* be the following text

*One day, his horse ran away quietly. The neighbors came to express their concern: "Oh, that's too bad. How are you going to work the fields now?" The farmer replied: "Good thing, Bad thing, Who knows?"*

*In a few days, his horse came back and brought another horse with her. Now, the neighbors were glad: "Oh, how lucky! Now you can do twice as much work as before!" The farmer replied: "Good thing, Bad thing, Who knows?"*

This variable is populated in the assignment template.

Complete the following of the exercises:

1. Wh-words in English are used in questions, relative clauses and exclamations. Consider the set of wh-words to consist exactly of the following members: *what, where, when, which, who, whom, whose, why.* 
   1. Report how many wh-words occurred in text2.
   2. Repeat the exercise for text7. Report how many wh-words occurred in text 7.
2. exercise 29
3. exercise 30. In this question, consider SimpleText for reporting your results.
4. exercise 40. Section 3.8 in “Sentence Segmentaion” lists an example of using Punkt. Use nltk.word\_tokenize() function to tokenize given text into words.
   1. Test your code on “ABC rural News”. Command nltk.corpus.abc.raw('rural.txt') allows one to access “ABC rural News” as a string. Report your results on this text.
   2. Report what your code computes for “ABC Science news” from ntlk.corpus.abc .

Submit

* printed solutions to the assignment on a due date in the beginning of the class.
* file nltkChapter3b\_template.py (populated with your coding solutions) via Canvas 30 minutes before the class on a due date.