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

### *Complete work individually.*

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

nltkChapter3a\_template.py

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

Complete the following of the exercises:

1. exercise 6 b,c,f: When you test your answers use the following text:

*One day, his horse ran away. 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?"*

In the rest of this assignment the text quoted above will be referred to as S*impleText*

1. exercise 7 a: In this question, consider SimpleText for testing your function.
2. exercise 21: In this question, use URL <https://www.cs.utexas.edu/~vl/notes/dijkstra.html>

What kind of differences do you see in the tokens you retrieve if you use nltk.word\_tokenize instead of re.findall(). Use nltk.WordNetLemmatizer() on your tokens before checking them against the Words Corpus. Do you find fewer or more unknown words? Why?

1. exercise 25: In this question, use SimpleText for reporting the outcome.

Submit

* printed solutions to the assignment on a due date in the beginning of the class.
* file nltkChapter3a\_template.py (populated with your coding solutions) via Canvas 30 minutes before the class on a due date.