### Assignment on Chapter 7 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 7 posted at:

<http://www.nltk.org/book/ch07.html>

On the Canvas, a file named

nltkChapter7\_template.py

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

Complete the following of the exercises:

1. The IOB format categorizes tagged tokens as I, O and B. Why are three tags necessary? What problem would be caused if we used I and O tags exclusively?

Provide a sentence to illustrate your point.

1. Write tag patterns to match noun phrases containing plural head nouns, e.g. "many/JJ researchers/NNS", "two/CD weeks/NNS", "both/DT new/JJ positions/NNS". Extend the “grammar” defined in Example 3 (Example 2.2 named code\_chunkex in the chapter 7).

by your regular expressions. Redefine the “cp” object from this example to use your new grammar. Use this object to parse sentenceSample defined as follows

sentenceSample = [("Many", "JJ"), ("little", "JJ"), ("dogs", "NNS"), ("barked", "VBD"), ("at", "IN"), ("cats", "NNS")]

Report your outcome.

1. Carry out the following evaluation tasks for the chunker you have developed in question 2.
2. Evaluate your chunker on first 100 sentences from a chunked corpus nltk.corpus.conll2000, and report the precision, recall and F-measure.
3. Compare the performance of your chunker to the baseline chunker discussed in the evaluation section of 3 (the very first chunker that does nothing).
4. Extend the “grammar” of your chunker by at least one more regular expression. Give rationally behind your extension. See whether this extension allows you to boost the performance of your chunker. Evaluate your new chunker on 100 sentences from a chunked corpus nltk.corpus.conll2000, and report the precision, recall and F-measure.

Submit

* printed solutions to the assignment on a due date in the beginning of the class.
* file nltkChapter7\_template.py (populated with your coding solutions) via Canvas 30 minutes before the class on a due date.