Assignment 3
Corpus linguistics/Spell checking
Due Wednesday, September 29

1. Do question #11 in chapter 3 of the draft textbook (p. 130).

2. Go to WordNet (http://wordnetweb.princeton.edu/perl/webwn) and look up time.
   (a) Focusing on the nouns, do you agree with the way time has been split into 10
       meanings? Or, should some of them be put together? Pulled apart? Are there
       any meanings which are missing?
   (b) Go to Project Gutenberg (http://www.gutenberg.org/) and download a book
       in a text format (in English). Verify that there are at least 10 occurrences of the
       word time in this book. You can use multiple books if need be.
   (c) For the first 10 instances (or more, if you want) of time, assign them the ap-
       propriate meaning: n1, ..., n10 for the nouns, v1, ..., v5 for the verbs. In your
       write-up, include enough context so I can see whether I agree or not.
   (d) Take one of the meanings with multiple occurrences: is there a common pattern
       in the surrounding words? Think of a regular expression that would help you to
       identify this meaning in new text.
   (e) Bonus: Use egrep on a new book, to see if your regular expression actually finds
       the meaning you want.

3. Do question #1a in chapter 2 of the draft textbook (p. 84–85).

4. Do question #2a in chapter 2 of the draft textbook (p. 85).

5.  • Make sure you have access to a computer that has Python.
    • Write a short program in Python which takes a string variable, calculates its
      length, and then prints out one letter per line (hint: string slices). In this case,
      your variable will contain the string "Hello world", but we might want to change
      that in the future, so be sure to use a while loop.

      Your output should look like this:

      H
      e
      l
      l
      o
      w
      o
      r
      l
      d

.