

Methods in Computational Linguistics 2  
LING 78100  
Fall 2010

Homework #2 - Programming in Python  
Due at 11:50 on October 8

1. Short Problems -- Can be answered in the interpreter

- A) [10] Review the discussion of looping with conditions in Section 1.4. Use a combination of `for` and `if` statements to loop over the words of the movie script for *Monty Python and the Holy Grail* (`text6`) and `print` all the uppercase words, one per line.
- B) [10] Write code that finds all words that occur at least 3 times in the Brown Corpus (The set of *all* words in the Brown Corpus can be found by `nlTK.corpus.brown.words()`)
- C) [10] Write a conditional statement that evaluates to true if a string starts with the word "a" or "the", contains two space characters, and either has 10 characters or ends with punctuation.

2. Open ended Programming Assignment -- Sentiment Analysis.

This question should be answered using stored python files.

- A) [5] Write two lists of positive and negative connoting words, like "good", "great", and "bad", "terrible".
- B) [10] Write a function that determines if a new word is positive, negative or neither. This may look something like the following.

```
def polarity(word):  
    if ... :  
        return "Positive"  
    elif ... :  
        return "Negative"  
    else:  
        return "Neither"
```

- C) [10] Write a function that determines if a sentence (represented as a list of words) is mostly positive or mostly negative. ("Mostly positive" sentences contain more positive words than negative words. "Mostly negative" sentences contain more negative words than positive words.)
- D) [10] We've said that negation is a big problem for sentiment analysis (as well as for automated semantic processing). Write a function that attempts to identify negation

in a sentence. One way to address this question is to identify the presence of common negation words, for example, “not”, “isn’t”, etc.

- E) [10] Include the function you wrote in D to your function that determines if a sentence is mostly positive or mostly negative (from C). If a negation is found in a sentence, you should change your determination from “mostly positive” to “mostly negative” and vice versa
- F) [10] Try your program out on a number of example sentences. Identify some sentences for which your program correctly determines the sentiment of the sentence. Identify some other sentences for which your program fails to determine the sentiment of the sentence.
- G) [15] Discuss the errors that your program makes. What makes these cases difficult to address? How might you address them with external language resources, or more complicated processing?