Activity 1

Oct. 7, 2010

Text Exercises

Let's practice some cool things you can do with text files in Python! Use your notes from the lecture slides to help you and consult a TA if you're confused. Whatever you don't finish here will become homework so don't worry about not finishing by the end of class. You will be writing code in a Python module. Be sure to save the Python file as 'YOURNAMEclass3Python.py'.

1. Download SAMPLEthreelives.txt from the course website and save it in your course folder. *Three Lives* is Gertrude Stein's first published work. Published in 1909, the book is separated into three stories, "The Good Anna," "Melanctha," and "The Gentle Lena."

2. Open up a python shell in IDLE and then open up a Python module.

3. Using what you learned in lecture, write code that 'opens' the file. Now our text is some value but in order to do something with it we need a string. Use the read() function to convert the text into a string, and give your text the name 'ThreeLivesText'. Print the first 75 characters of the text. Run your code to check your work so far.

4. Since we want to work with the words of the text we need to convert the really long string into a list of words. We can do this by using the split() function, which is used on strings and outputs a list. For example, assign `x = 'I love computer science.'` and then tell Python to `print x.split()`. Choose a variable (such as `wordlist`, for example), and assign it to the split version of 'ThreeLivesText'.

5. Let's say we want to make all the letters in the text lower case. We can do that using the lower() method! This method only works on a string but what we have now is a list. What we need to do is go through each element of the list (i.e. iterate over it) and call the lower() method on each string. There is one tricky thing we need to do in the iteration, however. The lower() method doesn't actually change the string; instead, it makes a lowercase copy of the old string, and returns the copy without altering the old string. Therefore, if we simply iterated over the list calling lower() on each string, the list

afterwards would not be lowercased because none of the strings would actually have been changed. Talk to a TA if you don't understand why this is so. The solution? Make a new list called `loweredlist` to hold the lowercase words, and add each lowercased string to it as you iterate through `wordlist`.

6. Write and test a python function that takes in a list and outputs another list which contains all the words from the original list, but lowercase.

7. How might you produce a list of all the words in the text without repeating any? You need to unique-ify the list! We wrote a function in lecture that will do this for us. What did we need to do first for it to work? Use what you learned in lecture to unique-ify the list and print the first 20 unique words in the text.

8. Can you think of cases when having a unique list of words in a text might be useful?

9. Print out the original text and underline all of the proper nouns by hand.

10. Reflect on how you carried out the previous task. Write down, in English, how you would find all the proper nouns in a text. This need not be more than a couple of sentences, but come to class prepared to discuss your ideas.

Handin

E-mail `YOURNAMEclass3Python.py` to `cs0931tas@cs.brown.edu` by Oct. 12, 2010 at 2:25 pm.