

CSCI 931 Class Activity October 21, 2010

We will start this activity in class. If you don't finish it in the time provided, please complete it for homework.

1. Create a folder on your desktop called *ConcActivity*. Copy into it the file *concordance.py* from the course web page for today's activity. Also copy *MobyDickStripped.txt* into *ConcActivity*. Start up IDLE. From the *File* menu of the Python shell, open the file *concordance.py* (within *Desktop/DictActivity*). Fix the default second parameter of the *getbook* function in *concordance* so that it refers to your *ConcActivity* directory.
2. The function *concordance_index* returns an index for the book supplied as its argument (given as a list of words). The function *concordance_print*, given a word, the book, and the book's index, prints out a list of all occurrences of that word along with their contexts. However, it doesn't indicate which chapters the occurrences are from. We'd like to fix that.

Create a new file, *MDconcordance.py*, in which you'll write some additional Python code that provides this functionality by building on top of what is provided by *concordance.py*. Here's a suggestion for how you might structure *MDconcordance.py*.

- a. It should import *concordance* so that it can build on top of it. (It may need to import *re* as well.)
- b. It should get the text of *Moby Dick* (using *getbook* from *concordance*). Note that what *getbook* returns is one long string.
- c. It should split the book into a list of chapters (each list element is a string containing one chapter of the book). Use *re.split* to do this: look at the book using *wordpad* and see if you can identify a pattern to use for splitting the text into chapters.
- d. Iterate through this list and create a new one, a list of lists, in which each top-level list is a list of words from a chapter. In other words, if this list is called *MDchapters*, then *MDchapters[i]* is the list of the words in chapter *i*.
- e. Using *concordance_index* (from *concordance*), create a separate index for each chapter (put all these indexes into yet another list: *MDindex*).
- f. *MDchapters* and *MDindex* contain the data you need to print a concordance. Define a function, *MDconcordance_print*, that takes a word as an argument and walks through the chapters of *Moby Dick*: if the word is in the chapter's index, then print out the chapter's title and call *concordance's concordance_print* function to print out the concordance for that word in that chapter.