

# Concordance

This assignment is worth 50 points (5% of the course grade) and must be completed and turned in **before 11:59 on Monday, October 31, 2011**. (HAPPY HALOWEEN)

## Assignment Overview

This assignment will give you more experience on the use of:

1. Dictionaries
2. File I/O
3. Functions
4. String manipulation

## Background

A concordance is a compendium of the words in a particular text, along with their local context. By local context, we mean the surrounding words that occurred in the sentence with the word. A concordance has a number of uses, including searching for idioms and phrases, identification of authors, and translations of partial texts.

This ([www.opensourceshakespeare.org/concordance](http://www.opensourceshakespeare.org/concordance)) is a shakespeare concordance. You can select any word, using its first letter, find how many times it was used, what works it occurred in and the local context of words it occurred in. For example, using the word "halcyon", you get the following:



OpenSourceShakespeare  
www.opensourceshakespeare.org

### Search results

1-2 of 2 total

KEYWORD: halcyon

- Double-click on any word to look it up in the concordance.
- For an explanation of each column, hover your cursor over the column's title.

#	Work	Character	Line	Text
1	<a href="#">Henry VI, Part I</a> [1, 2]	<a href="#">Joan la Pucelle</a>	328	Assign'd am I to be the English scourge. This night the siege assuredly I'll raise; Expect Saint Martin's summer, <b>halcyon</b> days, Since I have entered into these wars. Glory is like a circle in the water, Which never ceaseth to enlarge itself Till by broad spreading it disperse to nought. With Henry's death the English circle ends; Dispersed are the glories it included. Now am I like that proud insulting ship Which Caesar and his fortune bare at once.
2	<a href="#">King Lear</a> [11, 2]	<a href="#">Earl of Kent</a>	1139	That such a slave as this should wear a sword, Who wears no honesty. Such smiling rogues as these, Like rats, oft bite the holy cords atwain Which are too intrinse t' unloose; smooth every passion That in the natures of their lords rebel, Bring oil to fire, snow to their colder moods; Renege, affirm, and turn their <b>halcyon</b> beaks With every gale and vary of their masters, Knowing naught (like dogs) but following. A plague upon your epileptic visage! Smile you my speeches, as I were a fool? Goose, an I had you upon Sarum Plain, I'd drive ye cackling home to Camelot.

Program code and database © 2003-2011 George Mason University.  
All texts are public domain.

It identifies the word it occurred in, the line it occurred on, and the local context of

lines. That is a pretty big local context; we are going to aim a little smaller.

Because concordances are rather expensive (time wise) to develop, common words such as: and, a, the, those, him, her, etc. were ignored. These were called "stop words", and stop words are ignored in creating the concordance. However, they are *included in the local context*.

### Goals

Your goals are to prompt the user for a text file to analyze. You will also be provided with a file of stop words—words that will be ignored when you analyze the file. After creating the concordance, print the concordance results as indicated in the example output below. For each word in the concordance you will print:

- the word
- the total number of occurrences
- for each occurrence:
  - the line number in the file (the first line is line #1)
  - the local context. The local context is the original line with the specified word in all upper case.

### Project Specification

1. The stop word file is provided: english.stop.txt
2. Prompt the user for the file to analyze
3. Compute the concordance.
4. Print the contents of the concordance as shown below
5. Your program must use *two functions* of your own design. It is up to you what the functions do but you must have two functions.

### Deliverable

Turn in a single file named proj07.py which does all of the above.

### List of Files to Download

english.stop.txt, itsy\_bitsy\_spider.txt

### Notes and Hints

- for an open file object, `file_obj.readlines()` will return a **list**, where each element is a string, one of the lines of the file (a line ends in carriage return line feed, right!)
- Remember that case should not matter here. The string "Spider" and "spider" should both count as the same word.
- Punctuation should be stripped so that "spider." and "spider" both count as the same word.
- The context line printed should contain its original capitalization (except for the specified word that is all upper case) and original punctuation, e.g. notice

the explanation point in the example below.

**Example Output:**

Analyze what file: itsy\_bitsy\_spider.txt

Concordance for file itsy\_bitsy\_spider.txt

**itsy : Total Count: 2**

**Line:1: The ITSY Bitsy spider crawled up the water spout**

**Line:4: and the ITSY Bitsy spider went up the spout again!**

**crawled : Total Count: 1**

**Line:1: The Itsy Bitsy spider CRAWLED up the water spout**

**spout : Total Count: 2**

**Line:1: The Itsy Bitsy spider crawled up the water SPOUT**

**Line:4: and the Itsy Bitsy spider went up the SPOUT again!**

**washed : Total Count: 1**

**Line:2: Down came the rain and WASHED the spider out**

**spider : Total Count: 3**

**Line:1: The Itsy Bitsy SPIDER crawled up the water spout**

**Line:2: Down came the rain and washed the SPIDER out**

**Line:4: and the Itsy Bitsy SPIDER went up the spout again!**

**rain : Total Count: 2**

**Line:2: Down came the RAIN and washed the spider out**

**Line:3: Out came the sun and dried up all the RAIN**

**water : Total Count: 1**

**Line:1: The Itsy Bitsy spider crawled up the WATER spout**

**bitsy : Total Count: 2**

**Line:1: The Itsy BITSY spider crawled up the water spout**

**Line:4: and the Itsy BITSY spider went up the spout again!**

**dried : Total Count: 1**

**Line:3: Out came the sun and DRIED up all the rain**

**sun : Total Count: 1**

**Line:3: Out came the SUN and dried up all the rain**