CS210A2

CS 210
West Virginia University
fall semester 2008
August 19, 2008
Assignment 1
due Tuesday, September 30, 2008

Overview: Write a Python program that constructs a concordance.

Input: a text document.

Output: a text document with one section per word that appears in the original document. The section for a word consists of:

    the word,
    the number of times the word appears in the original document
    a list of all lines (with line number) in which the word appears
    The sections are arranged in alphabetical order by word.

Example input:

```
1 THROUGH THE LOOKING-GLASS
2
3 by LEWIS CARROLL
4
5 THE MILLENNIUM FULCRUM EDITION 1.7
6
7
8
9
10 CHAPTER 1
11
12 Looking-Glass house
13
14
15 One thing was certain, that the WHITE kitten had had nothing to
16 do with it:--it was the black kitten's fault entirely. For the
17 white kitten had been having its face washed by the old cat for
18 the last quarter of an hour (and bearing it pretty well,
19 considering); so you see that it COULDN'T have had any hand in
20 the mischief.
```

. . .

Example output:

```
. . .
it 4
16 do with it:--it was the black kitten's fault entirely. For the
18 the last quarter of an hour (and bearing it pretty well,
19 considering); so you see that it COULDN'T have had any hand in

its 1
17 white kitten had been having its face washed by the old cat for

kitten 2
15 One thing was certain, that the WHITE kitten had had nothing to
17 white kitten had been having its face washed by the old cat for

kitten's 1
16 do with it:--it was the black kitten's fault entirely. For the

. . .
white 2
15 One thing was certain, that the WHITE kitten had had nothing to
17 white kitten had been having its face washed by the old cat for
```

Additional comments:

1. For text to test your program, go to the Project Gutenberg site and download "Through the Looking Glass."
http://www.gutenberg.org/etext/12
Delete the header and footer of your copy of the file before using it as input for your program.

2. You may either have your program compute line numbers for the text or may edit the file to add line numbers to it.
If the text is stored in file
lookingglass
an easy way to add line numbers in *nix is

```
cat -n <lookingglass >lookingglass2
```

3. Your program should recognize all casings of a word as being the same but when displaying the line in which the word occurs should show the original casing. (Notice "WHITE" and "white" in the sample output above.)

4. For more information on what a concordance is, see the Wikipedia article on "Concordance (publishing).

http://en.wikipedia.org/wiki/Concordance_(publishing)