CS 101
Fall 2012
Makeup program
Due Sunday night, December 9

This optional program can replace your lowest score on any one of the first 8 programs.

The basic problem description can be found on page 466 of your text.

You are given a text file (debate2.txt) containing a transcript of the second presidential debate between Barack Obama and Mitt Romney. Your program will read through the file, filter out the 'stop words,' and produce an HTML tag cloud for each speaker.

Break this problem down into parts.

- In reading through the file, you must first separate the words by speaker. Note that every change of speaker is clearly indicated.
- You then need to put the text into consistent text (all lower-case), filter out certain names and titles (“Crowley”, “Romney”, “Obama”, “President”, “Governor”, “Candy”, “Mr.”) , and filter out punctuation.
- Next, filter out the commonly-occuring 'stop words' listed in the file stop.txt.
- Then, for each speaker, determine how many times each word was used. (A dict will be useful here—the keys are words, the values are the counts of how many times that word was used.)
- Based on the count, determine the font size to be used for each word. You may need to experiment a bit until you get something that looks right. Words that appeared most frequently should be in larger sizes than words appearing less frequently.
- Generate the HTML tag for each word based on its font size. You are given a function MakeHTMLWord() to do this; it takes a string and a font size as its parameters, and returns (as a string) the HTML code necessary to print that string in that font size.
- Once you have the HTML tags for each word, build your tags up into a single string. You then need to put that string into formatted HTML. Again, you are given a function for this, MakeHTMLBox(). It takes a string of HTML as its parameter, and returns a new string containing the HTML code to print the string in a formatted box. You will take the string returned by MakeHTMLBox() and save it to an HTML file (using the function printHTMLfile(), also provided). You should then be able to open the HTML file in any web browser.

```
>>> ================================ RESTART ================================
>>>
BidenCounts =  [('afghanistan', 12), ('america', 13), ('back', 11), ('barack', 40), ('billion'
, 14), ('care', 11), ('change', 18), ('class', 12), ('economic', 12), ('fact', 15), ('george',
10), ('give', 11), ('governor', 17), ('health', 11), ('iraq', 10), ('john', 70), ('making', 12
), ('mccain', 52), ('middle', 13), ('money', 10), ('nuclear', 10), ('number', 15), ('obama', 3
3), ('people', 24), ('plan', 13), ('policy', 23), ('president', 18), ('senate', 10), ('states'
, 16), ('street', 10), ('support', 17), ('talk', 12), ('tax', 27), ('time', 12), ('troops', 10
), ('understand', 12), ('united', 16), ('voted', 27), ('war', 17), ('years', 16)]
PalinCounts =  [('administration', 12), ('afghanistan', 16), ('america', 16), ('american', 19)
, ('americans', 15), ('back', 10), ('barack', 22), ('change', 12), ('economy', 13), ('energy',
29), ('good', 17), ('government', 18), ('governor', 13), ('iraq', 13), ('john', 34), ('make',
12), ('mccain', 35), ('nuclear', 11), ('obama', 19), ('people', 29), ('plan', 16), ('put', 11)
, ('putting', 11), ('reform', 12), ('senator', 13), ('state', 15), ('talk', 12), ('tax', 19),
('taxes', 17), ('thing', 10), ('ticket', 10), ('time', 11), ('troops', 10), ('voted', 13), ('w
ar', 13), ('work', 16), ('years', 12)]
>>>
```