Project Overview:

The program I wrote for hw3 was basically something that can tell you whether if a book can be read by a 5th grade. The program takes in a URL of a book from ProjectGutenburg, and will basically compare each word of a book to another list of words that 5th graders are meant to know. The more words that match, the more likely a 5th grader will be able to read them.

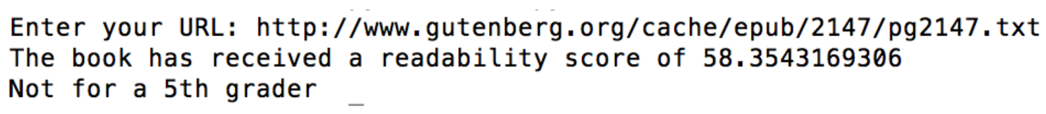
Implementation:

So the only things required for the program is the main.py which is code, and then the word2.txt which is the file with the 5th grade vocab. I only used the pattern library to download the contents of the URL, and the string library to manipulate the text so that it is easier to compare. To start the program just run main.py with the words2.txt file in the same folder and all should be well. I worked a lot with lists to compare the data.

NOTE: nwords.txt will be created after the program is run. It contains all the words that were not in the words2.txt file. This is so that I can add words later on.

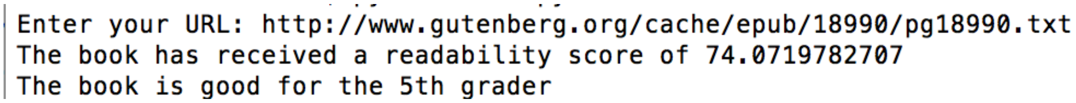
Output:

Image shows that the URL is not for a 5th grader



```
Enter your URL: http://www.gutenberg.org/cache/epub/2147/pg2147.txt
The book has received a readability score of 58.3543169306
Not for a 5th grader  _
```

Image shows that it is.



```
Enter your URL: http://www.gutenberg.org/cache/epub/18990/pg18990.txt
The book has received a readability score of 74.0719782707
The book is good for the 5th grader
```

The readability score is basically the percentage of the similarity of words. So 58 is pretty bad. Generally 70 plus is what I get for children’s books from Gutenburg.

Reflection:

I think overall the thing that I wanted to make was not made. Instead I have this bare program that could use a lot of tweaking and upgrading. I think that the words list can be improved. More things such as length of words, and sentences can be taken into account to give better results. I could also probably break up the analyze function into smaller pieces. Definitely put in something that can reads books from a lot of sites. The thing that I know worked well was the extra file that was created which contained words which were not in the list. It could be improved with some sort of sorter though.