FILE ANALYSIS ASSIGNMENT

This assignment will examine features of a text file.

Find the following

1. The total number of words in the file.
2. The total number of unique words in the file.
3. An alphabetical listing of the words.
4. The top 30 words which occur the most frequently, sorted in descending order by frequency.
5. A list of 1 letter words, 2 letter words, 3 letter words, etc
6. Longest word list in alphabetical order.
7. A list of misspelled words based on words in words.txt
8. The Grade Level Readability Index

**Grade Level Readability Formula**   
  
**Step 1:** Calculate the average number of words used per sentence.   
  
**Step 2:** Calculate the average number of syllables per word.   
  
**Step 3:** Multiply the average number of words by 0.39 and add it to the average number of syllables per word multiplied by 11.8.   
  
**Step 4:** Subtract 15.59 from the result.   
  
The specific mathematical formula is:   
  
GLRA = (0.39 x ASL) + (11.8 x ASW) - 15.59   
  
Where,   
  
GLRA = Grade Level Reading Age   
  
ASL = Average Sentence Length (i.e., the number of words divided by the number of sentences)   
  
ASW = Average number of Syllable per Word (i.e., the number of syllables divided by the number of words)

* A word is any sequence of characters delimited by white space, whether or not it is an actual English word.
* A syllable is defined by the following rules:

Each group of adjacent vowels (a, e, i, o, u, y) counts as one syllable (for example,

the “ea” in “real” contributes one syllable, but the “e . . . a” in “regal”

count as two syllables). However, an “e” at the end of a word doesn’t count as

a syllable. Also, each word has at least one syllable, even if the previous rules

give a count of 0.

* A sentence is ended by a period, colon, semicolon, question mark, or exclamation mark.

Analyzing the results is a simple exercise. For instance, a score of 5.0 indicates a grade-school level; i.e., a score of 9.3 means that a ninth grader would be able to read the document. This score makes it easier for teachers, parents, librarians, and others to judge the readability level of various books and texts for the students.