**Lab 1410 - Flesch**

Description

For this assignment, you are to write a program that computes the Flesch Readability Score for a document. The Flesch Readability Score is computed using a formula developed iin the early 1940s. It measures the average sentence length in words and the average word length in syllables. Put these two numbers into a formula and get a number between 0 and 100 that shows you the difficulty of a piece of writing.

Here is the algorithm according to the original author, Rudolf Flesch, along with simplifications you should make for this project:

* "Count the words in your piece of writing. Count as single words contractions, hyphenated words, abbreviations, figures, symbols and their combinations." That is not always easy to do accurately, so if you are using the *chn* utility package, it is sufficient to read tokens and count each as a word.
* "Count the syllables in your piece of writing. Count the syllables in words as they are pronounced. Count abbreviations, figures, symbols and their combinations as one-syllable words." A syllable needs a vowel, so each vowel (a,e,i,o,u,y) or group of vowels counts as a syllable. However, don't count an "e" at the end of a word as a syllable (as in the word "here"). Every word must be pronounced, so there will be at least one syllable even if the rules say there are none.
* "Count the sentences in your piece of writing. Count as a sentence each full unit of speech marked off by a period, colon, semicolon, dash, question mark or exclamation point. Disregard paragraph breaks, colons, semicolons, dashes or initial capitals within a sentence." For this assignment, ignore dashes entirely and just count as a sentence text that ends in a period, colon, semicolon, question mark or exclamation point.
* "Figure the average number of syllables per word. Divide the number of syllables by the number of words."
* "Figure the average number of words per sentence. Divide the number of words by the number of sentences."
* Find the readability score. Use Flesch's formula: "Multiply the average sentence length by 1.015. Multiply the average word length by 84.6. Add the two numbers. Subtract this sum from 206.835. The balance is your readability score."

The readability score is a number, usually between 0 and 100, indicating how difficult the text is to read. Higher numbers indicate easier reading material.

Examples

Flesch has provided [some examples](http://www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.html) of readability score calculations from various publications:

|  |  |
| --- | --- |
| Comics | 92 |
| *Seventeen* | 67 |
| *Time Magazine* | 52 |
| standard auto insurance policy | 10 |
| Internal Revenue Code | -6 |

The Flesch Readability Score can be translated into educational levels using this table:

|  |  |
| --- | --- |
| 91-100 | 5th grade |
| 81-90 | 6th grade |
| 71-80 | 7th grade |
| 61-70 | 8th and 9th grade |
| 51-60 | 10th to 12th grade |
| 31-50 | college student |
| 0-30 | college graduate |

Assignment

Given a source file, your program should compute the Flesch Readability Score. It must also provide detailed statistics through accessor methods. Here is the test program:

public class TestFlesch

{

public static void main (String[] args)

{

Flesch ms = new Flesch("Gettysburg.txt");

System.out.println( "Source text file: " + ms.getFilename() ) ;

System.out.println( "Readability score: " + ms.getLegibilityIndex() ) ;

System.out.println( "Educational level: " + ms.getEducationalLevel() );

System.out.println( "Syllables: " + ms.getNumSyllables() );

System.out.println( "Words: " + ms.getNumWords() );

System.out.println( "Sentences: " + ms.getNumSentences() );

}

}

Here is the format of the output using the Gettysburg Address:

Source text file: Gettysburg.txt

Readability score: 61.0

Educational level: 8th and 9th grade

Syllables: 374

Words: 265

Sentences: 10

Several files are provided for testing:  
[test-6th-grader.txt](http://rfrank.net/r/cslabs-final/1410-flesch/test-6th-grader.txt) suitable for a sixth grader, legibility index 87.  
[test-college-grad.txt](http://rfrank.net/r/cslabs-final/1410-flesch/test-college-grad.txt) suitable for a college graduate, legibility index 28.  
[Gettysburg.txt](http://rfrank.net/r/cslabs-final/1410-flesch/Gettysburg.txt) Gettysburg Address.

You may want to test your code with other text files before submitting you code, *Flesch.java* for grading. I will test it with very easy and very difficult passages and will expect reasonable results.

|  |  |
| --- | --- |
| Copyright © 2010 by Asylum Computer Services LLC | Return to [CS Labs](http://rfrank.net/r/cslabs-final/cslabs.html) |