Use Flesch Score to Measure Readability

Goals:

- Practice with Java String operations
- Learn how to write complicated regular expression
- Implement polymorphism with override

Two starter files are provided: Document.java and BasicDocument.java. Please follow the guideline below to complete the missing implementations in those two classes. Note that, the Document class is an abstract class, which is entended by the BasicDocument class.

Your Tasks:

1. Implement countSyllables, getNumSentences, getNumWords and getNumSyllables methods following the instructions found in the comments of Document.java and BasicDocument.java.

Several test cases are provided in the main method as well as a helper method you can use to write your own tests. To get full credits, you need to pass all the test cases. The key is to follow the exact definitions of what constitutes a syllable, word and sentence given in the comments.

A helper method called getTokens is provided in the Document class, which can take a regular expression and return a list of matched “tokens” from the string member variable. You may leverage this method to implement the required methods. It is possible to write regex's to count the number of syllables directly, but this is not necessarily the best approach. If you find yourself doing mental gymnastics to come up with a single regex to count syllables directly, that's usually an indication that there's a simpler solution (i.e., using a loop to iterate the characters for counting). Just because a piece of code (e.g. a regex) is shorter does not mean it is always better.

Under the rules for what defines a syllable, the words “the”, “fly”, “yes”, “cave” and “double” all have 1 syllable, but "segue" has two syllables. Notice that this isn’t exactly correct (“double” actually has 2 syllables), but it’s close enough for our purposes. Here are some more examples with the number of syllables your method should return to help you: "contiguous" (3 syllables), "sleepy" (2 syllables), "obvious" (2 syllables), "toga" (2 syllables). Our rules get a lot wrong, especially when you have more than 2 vowels in a row, but these are the rules we will test you against.

2. Implement the getFleschScore method in Document.java to calculate the Flesch Score of the text in the document. You should use the following formula, and make calls to the getNumSyllables, getNumWords, and getNumSentences you just implemented.

$$Flesch\ score = 206.835 - 1.015\left(\frac{\#\ words}{\#\ sentences}\right) - 84.6\left(\frac{\#\ syllables}{\#\ words}\right)$$

You should test your code by calculating the Flesch score by hand on some very basic documents and then calling your method from the main method to make sure it's giving the correct output.

What and how to submit

Submit Document.java, BasicDocument.java and the screenshot of the test cases’ output.