

Readability test

The *Flesch Reading Ease* test is generally regarded as the most accurate score to express how quickly a particular English text can be read. Several U.S. government agencies use this test to check the readability of their documents. For example, the State of Florida requires a minimum *Flesch Reading Ease* score of 45 for life insurance contracts. The score was also built in into several popular word processors such as Microsoft Office Word and WordPerfect. The *Flesch Reading Ease* formula goes as follows:
$$FRES = 206.835 - 1.015 \times \frac{\text{total words}}{\text{total sentences}} - 84.6 \times \frac{\text{total syllables}}{\text{total words}}$$
 The result of the formula is the Flesch Reading Ease Score (FRES). The higher the FRES, the easier it is to understand the text. A score between 0 and 30 is considered to be readable for university graduates. A text with a score between 60 and 70 should be readable for pupils from 13-15 years, and a score between 90 and 100 should be readable for pupils at the end of primary school. The highest (easiest) readability score is about 120, this is achieved when each sentence is composed of two monosyllabic words. The score has no theoretical limit. It is always possible to reduce the score by adding words with many syllables.

Although the application of FRES seems simple, we still have to clearly formulate how sentences, words and syllables are counted. **Sentences** are separated by punctuation marks: periods(.), exclamation marks(!), question marks(?), colons(:) and semicolons (;). The fragment after the last punctuation mark should also be counted as a sentence, unless the fragment is empty or consists only of white space. Each group of consecutive letters (a-z, A-Z) forms a separate **word**. In English texts an apostrophe (') is exclusively used to indicate a contraction of two words. The fragment "it's" thus contains two words. To count **syllables** the following rules are applied:

- words of maximum three letters always consist of only one syllable
- each vowel (aeiouy) in a word is considered as one syllable, except in the following cases
- the *silent* e in word ends that are equal to -es, -ed or -e (except -le) are ignored (as if they weren't there)
- sequences of vowels count as one syllable

Assignment

The aim of this assignment is to determine the *Flesch Reading Ease* score of the text in a given text file. To do this, follow these steps.

- Write a function `syllables` that returns the number of syllables of a given word. The word must be passed to the function as an argument.
- Use the function `syllables` to write a function `words` to which a sentence should be passed as an argument. This function must determine the number of words, and the total number of syllables in all words of the sentence, and return these two values as a tuple.
- Use the function `words` to write a function `sentences` that file must be passed the file name of a text as an argument. The function must determine *i*) how many sentences occur in the text file *ii*) the total number of words in the text file, and *iii*) the total number of syllables in all words of the file. These three values must be returned by the function as a tuple.
- Write a function that `FRES` to which three arguments must be passed: the number of sentences in a text, the total number of words in the text, and the total number of syllables in

all words of the text together. The function should return the *Flesch Reading Ease* score of the text based on these three values.

Example

In the example session below the *Flesch Reading Ease* score is calculated of a couple sentences of chapter 64 from the classic [Moby Dick](#) by Herman Melville. These sentences can be found in the text file [moby_dick.txt](#).

```
>>> syllables('reliable')
3
>>> syllables('creature')
2
>>> syllables('three')
1

>>> words('The Australian platypus is seemingly a hybrid of a mammal and reptilian creature.')
(13, 24)
>>> words("Stubb's whale had been killed some distance from the ship.")
(11, 13)

>>> sentences('moby_dick.txt')
(12, 220, 298)

>>> FRES(12, 220, 298)
73.632121212123
```

De *Flesch Reading Ease* test wordt algemeen beschouwd als de meest nauwkeurige score om uit te drukken hoe vlot een bepaalde Engelstalige tekst kan gelezen worden. Verschillende Amerikaanse overheidsinstanties gebruiken deze test om de leesbaarheid van hun documenten te controleren. Zo eist de staat Florida dat contracten voor levensverzekeringen een minimale *Flesch Reading Ease* score van 45 hebben. De score werd dan ook ingebouwd in verschillende populaire tekstverwerkers zoals Microsoft Office Word en WordPerfect. De *Flesch Reading Ease* formule luidt als volgt:
$$FRES = 206.835 - 1.015 \times \frac{\text{aantal woorden}}{\text{aantal zinnen}} - 84.6 \times \frac{\text{aantal lettergrepen}}{\text{aantal woorden}}$$
 Het resultaat van de formule is de *Flesch Reading Ease* score (FRES). Hoe hoger de FRES, hoe gemakkelijker het is om de tekst te begrijpen. Een score tussen 0 en 30 wordt beschouwd als leesbaar voor universitair geschoolden. Een tekst met een score tussen 60 en 70 zou leesbaar moeten zijn voor scholieren van 13-15 jaar, en een score tussen 90 en 100 zou leesbaar moeten zijn voor leerlingen aan het eind van de lagere school. De hoogst mogelijke (gemakkelijkste) leesbaarheidsscore is ongeveer 120. Deze wordt bekomen wanneer elke zin bestaat uit twee éénlettergrepige woorden. De score heeft geen theoretische ondergrens. Het is steeds mogelijk om de score te verlagen door woorden met veel lettergrepen toe te voegen.

Hoewel de toepassing van de FRES eenvoudig lijkt, moeten we toch duidelijk formuleren hoe zinnen, woorden en lettergrepen geteld worden. **Zinnen** worden van elkaar gescheiden door leestekens: punten (.), uitroepingstekens (!), vraagtekens (?), dubbelepunten (:) en puntkomma's (;). Het fragment na het laatste leesteken moet dus ook als een zin geteld worden, tenzij dat fragment leeg is of enkel uit witruimte bestaat. Elke groep van opeenvolgende letters (a-z, A-Z) vormt een afzonderlijk **woord**. In Engelstalige teksten wordt een apostrofe (') uitsluitend gebruikt om de samentrekking van twee woorden aan te geven. Het fragment "It's" bevat dus twee woorden. Om **lettergrepen** te tellen worden de volgende regels gebruikt:

- woorden met hoogstens drie letters bestaan altijd uit één enkele lettergreep
- elke klinker (aeiouy) in een woord wordt beschouwd als één lettergreep, behalve in onderstaande gevallen
- de *silent e* in wordeindes die gelijk zijn aan -es, -ed of -e (behalve -le) wordt genegeerd (alsof ze niet in het woord staan)
- opeenvolgingen van klinkers tellen als één enkele lettergreep

Opgave

Het doel van deze opgave is om de *Flesch Reading Ease* score te bepalen van de tekst in een gegeven tekstbestand. Hiervoor ga je als volgt te werk.

- Schrijf een functie `lettergrepen` die het aantal lettergrepen van een gegeven woord teruggeeft. Het woord moet als argument aan de functie doorgegeven worden.
- Gebruik de functie `lettergrepen` om een functie `woorden` te schrijven waaraan een zin als argument moet doorgegeven worden. Deze functie moet het aantal woorden en het totaal aantal lettergrepen in alle woorden van de zin bepalen, en deze twee waarden als een tuple teruggeven.
- Gebruik de functie `woorden` om een functie `zinnen` te schrijven waaraan de bestandsnaam van een tekstbestand als argument moet doorgegeven worden. De functie moet bepalen *i)* hoeveel zinnen er in het tekstbestand voorkomen, *ii)* wat het totaal aantal woorden in het tekstbestand, en *iii)* wat het totaal aantal lettergrepen is in alle woorden van het bestand. Deze drie waarden moeten als een tuple door de functie teruggegeven worden.
- Schrijf een functie `FRES` waaraan drie argumenten moeten doorgegeven worden: het aantal zinnen in een tekst, het totaal aantal woorden dat in die tekst voorkomt, en het totaal aantal lettergrepen in alle woorden van de tekst samen. De functie moet op basis van deze drie waarden de *Flesch Reading Ease* score van de tekst berekenen en als resultaat teruggeven.

Voorbeeld

In onderstaande voorbeeldsessie wordt onder andere de *Flesch Reading Ease* score berekend voor een aantal zinnen uit hoofdstuk 64 van de klassieker [Moby Dick](#) van Herman Melville. Deze zinnen zijn terug te vinden in het tekstbestand [moby_dick.txt](#).

```
>>> lettergrepen('reliable')
```

```
3
```

```
>>> lettergrepen('creature')
```

```
2
```

```
>>> lettergrepen('three')
```

```
1
```

```
>>> woorden('The Australian platypus is seemingly a hybrid of a mammal and reptilian creature.')
```

```
(13, 24)
```

```
>>> woorden("Stubb's whale had been killed some distance from the ship.")
```

```
(11, 13)
```

```
>>> zinnen('moby_dick.txt')
```

```
(12, 220, 298)
```

```
>>> FRES(12, 220, 298)
```

```
73.632121212123
```

