**BFSU Readability Analyzer 3**

BFSU Readability Analyzer 3 is a tool designed to provide basic readability statistics for English texts. The tool was compiled from a Python script generated by Claude 3.5 Sonnet, using a prompt composed by Jiajin Xu of BFSU Corpus Research Group, the National Research Centre for Foreign Language Education, Beijing Foreign Studies University.

This tool can compute eleven lexical and readability scores, including Flesch Reading Ease, Text Difficulty (100 – Flesch Reading Ease), Flesch-Kincaid Grade Level, Sentence Count, Average Syllables per Word (ASW), Average Word Length (AWL), Token Count, Type Count, Lemma Count, Word STTR (Standardized Type-Token Ratio), and Lemma STTR (Standardized Type-Token Ratio).

Users can upload one or more English texts in plain text format (\*.txt) by clicking on the button [Choose Files]. For calculating STTR, users need to set the basis for the calculation, with the default being 100 words. This setting can be customized via a dropdown menu [STTR Basis], allowing adjustments in increments of 100 words, ranging from 100 to 2,000 words.

Users can click [Compute] to process the uploaded texts.

\*\*\*

**Understanding Readability Scores in the Readability Results**

**Flesch Reading Ease score**

This analysis scores text readability on a 100-point scale. Higher scores indicate easier comprehension. Most standard files typically score between 60 and 70.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Score mapping table** | | |
| **Flesch Reading  Ease Score** |  | **Readability Level** |
| 0 - 29 |  | Very difficult |
| 30 - 49 |  | Difficult |
| 50 - 59 |  | Fairly difficult |
| 60 - 69 |  | Standard |
| 70 - 79 |  | Fairly easy |
| 80 - 89 |  | Easy |
| 90 - 100 |  | Very easy |

The formula for the Flesch Reading Ease score is:

Flesch Reading Ease = 206.835 – (1.015\*ASL) – (84.6\*ASW)

where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

**Text Difficulty score**

To make the readability score based on the Flesch Reading Ease test easier to interpret, we inverted the scale so that a score of 0 represents the easiest text and 100 represents the most difficult. The following equation was applied to calculate the text difficulty score.

(Flesch Reading Ease based) Text Difficulty = 100 – Flesch Reading Ease score

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Score mapping table** | | |
| **Text Difficulty** |  | **Readability Level** |
| 0 - 29 |  | Very easy |
| 30 - 49 |  | Easy |
| 50 - 59 |  | Fairly easy |
| 60 - 69 |  | Standard |
| 70 - 79 |  | Fairly difficult |
| 80 - 89 |  | Difficult |
| 90 - 100 |  | Very difficult |

**Flesch-Kincaid Grade Level score**

This test rates text on a U.S. school grade level. For example, a score of 8.0 means that an eighth grader can understand the text. Most standard texts typically score between 7.0 to 8.0.

The formula for the Flesch-Kincaid Grade Level score is:

Flesch-Kincaid Grade Level = (.39\*ASL) + (11.8\*ASW) – 15.59

where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

**Sentences**

This metric indicates the total number of sentences extracted from the uploaded text(s).

**ASL** = Average Sentence Length

**AWL** = Average Word Length

**Tokens** = the total number of all occurrences of alphanumeric symbols

**Word Types** = the total number of distinct words (e.g., 10 occurrences of ‘do’ are counted as 1 type for the word ‘do’)

**Lemma Types** = the total number of base forms (e.g., ‘do’ is the lemma for do, does, did, doing, done if any).

**Standardised TTR (Word STTR)** (the following definition was adapted from WordSmith manual)

The standardized type/token ratio (STTR) is computed by dividing the number of unique words (types) by the total number of words (tokens) in a given text, providing a measure of lexical diversity that accounts for text length. By default, n = 100. In other words, the ratio is calculated for the first 100 running words, then calculated afresh for the next 100, and so on to the end of your text or corpus. A running average is computed, which means that you get an average type/token ratio based on consecutive 100-word chunks of text. Texts with less than 100 words (or whatever n is set to) will get a standardized type/token ratio of 0.

**Lemma STTR**

Lemma STTR calculates the ratio of the number of lemma types and the total number of lemmas, based on a specified chunk—100 words by default.

The results are automatically saved as **readability\_results.txt** in the source folder of the selected texts.

Mingchen Sun contributed to validating the results and optimizing the code.

© 2024 Jiajin Xu

National Research Centre for Foreign Language Education, Beijing Foreign Studies University

<https://corpus.bfsu.edu.cn>

Last edited: December 14, 2024