

Text Readability Analyzer Documentation

Kashif Khan(22P9005)

February 15, 2024

Abstract

This document provides documentation for the Text Readability Analyzer program, including its purpose, description, and usage.

1 Introduction

The Text Readability Analyzer program is designed to calculate the readability index of a set of text files within a specified directory. This index is based on the Automated Readability Index (ARI), which measures the complexity of text. The program accomplishes this by counting characters, words, and sentences in each text file, and then applying a predefined formula to calculate the readability index.

1.1 Formula

$$4.71(\text{characters}/\text{words}) + 0.5(\text{words}/\text{sentences})21.43$$

2 Program Description

The program utilizes a multi-pass algorithm to iterate through each text file in the specified directory. This algorithm includes the following features:

2.1 Count_Charac() Function

This function counts the number of characters in the input text. It iterates through each character, incrementing a count if the character is alphanumeric.

2.2 Count_Words() Function

This function counts the number of words in the input text. It identifies words based on whitespace and sentence terminations (e.g., periods, exclamation marks). A word is defined as any continuous sequence of alphabetic characters delimited by whitespace or sentence endings.

2.3 Count_Sent() Function

This function counts the number of sentences in the input text. It identifies sentence terminations and increments a count accordingly.

3 Main Execution

The program begins by specifying a directory containing text files. It then reads each file sequentially and utilizes the aforementioned functions to calculate the number of characters, words, and sentences in each file. Finally, it applies the predefined formula to calculate the readability index, rounding the result to the nearest integer.

4 Conclusion

The Text Readability Analyzer program provides a convenient method for assessing the readability of text files. By utilizing the ARI metric, it offers insights into the complexity of written content, which can be valuable for various applications such as educational material evaluation, content creation, and more.