

Assignment 1 Report

Tazmeen Afroz
22P-9252
BS-AI-4A

February 15, 2024

Abstract

Automated Readability Index(ARI) is a score designed to gauge the readability of a text, and can be used as a feature for developing models for text authorship. This document presents a Python program that calculates the ARI for a set of text files. The program reads the content of each file, counts the number of sentences, words, and characters, and then calculates the ARI.

1 Introduction

The purpose of this assignment is to get you started with Python programming, developing familiarity with Google Colab and its connectivity with Google Drive.

2 Program Description

2.1 Counting Sentences

The function `count_sentences(text)` is used to count the number of sentences in a given text. A sentence is defined as any sequence of characters ending with a sentence terminator. The sentence terminators considered in this program are a period (.), colon (:), semicolon (;), question mark (?), and exclamation mark (!).

The function iterates over each character in the text. If the character is a sentence terminator, it increments a counter (`sentence_count`). If the text does not contain any sentence terminators, the function assumes that the text consists of one sentence.

Here is the Python code for the function:

```
def count_sentences(text):  
    sentence_count = 0  
    for terminator in text:  
        if terminator in sentence_terminators:  
            sentence_count += 1  
    # If a text has no sentence characters assume it has 1 sentence.  
    return max(sentence_count, 1)
```

2.2 Counting Words

The function `count_words(text)` is used to count the number of words in a given text. A word is defined as any sequence of characters delimited by a word terminator. The word terminators considered in this program are a space, tab, newline, apostrophe ('), underscore (_), hyphen (-), comma (,), or sentence terminator.

The function iterates over each character in the text. If the character is a word terminator, it replaces the terminator with a space. This is done to ensure that words are correctly split in the next step. After all terminators have been replaced, the function splits the text into words using the `split()` method, which splits a string into a list where each word is a list item. The function then returns the number of items in the list, which is the number of words in the text.

Here is the Python code for the function:

```
def count_words(text):
    for terminator in words_terminator:
        text = text.replace(terminator, '-')
    words = text.split()
    return len(words)
```

2.3 Counting Characters

The function `count_characters(text)` is used to count the number of characters in a given text. A character is defined as any alphanumeric character, i.e., any letter or digit. The function iterates over each character in the text. If the character is alphanumeric, it increments a counter (`character_count`). The function then returns the value of `character_count`. Here is the Python code for the function:

```
def count_characters(text):
    character_count = 0
    for char in text:
        if char.isalpha() or char.isdigit():
            character_count += 1
    return character_count
```

2.4 Calculating Automated Readability Index

The function `calculate_ari(text)` is used to calculate the Automated Readability Index (ARI) for a given text. It first counts the number of sentences, words, and characters in the text using the previously defined functions. It then calculates the ARI using the formula $4.71 * (\text{characters} / \text{words}) + 0.5 * (\text{words} / \text{sentences}) - 21.43$. The result is rounded up to the nearest whole number using the `math.ceil` function. If the ARI is less than 0, it is set to 0. Here is the Python code for the function:

```
def calculate_ari(text):
    sentences = count_sentences(text)
    words = count_words(text)
    characters = count_characters(text)

    ari = math.ceil(4.71 * (characters / words) + 0.5 * (words / sentences) - 21.43)

    return max(ari, 0)
```

2.5 Printing the Result

The function `print_result(filename, text, sentences, words, characters, ari)` is used to print the results for a text file.

2.6 Returning File Paths

The function `return_filepath()` returns a list of all .txt files in a specified directory.

2.7 Testing Files

The `main()` function is the main function of the program. It gets the list of all text files, reads the content of each file, counts the number of sentences, words, and characters, calculates the ARI, and prints the results.

Here is the Python code for these functions:

```
def print_result(filename, text, sentences, words, characters, ari):
    print(f'Text-File-{filename}:-{text}')
    print(f'Number-of-sentences:-{sentences}')
```

```

print(f'Number of words: {words}')
print(f'Number of characters: {characters}')
print(f'Readability index: {ari}')
print('\n')

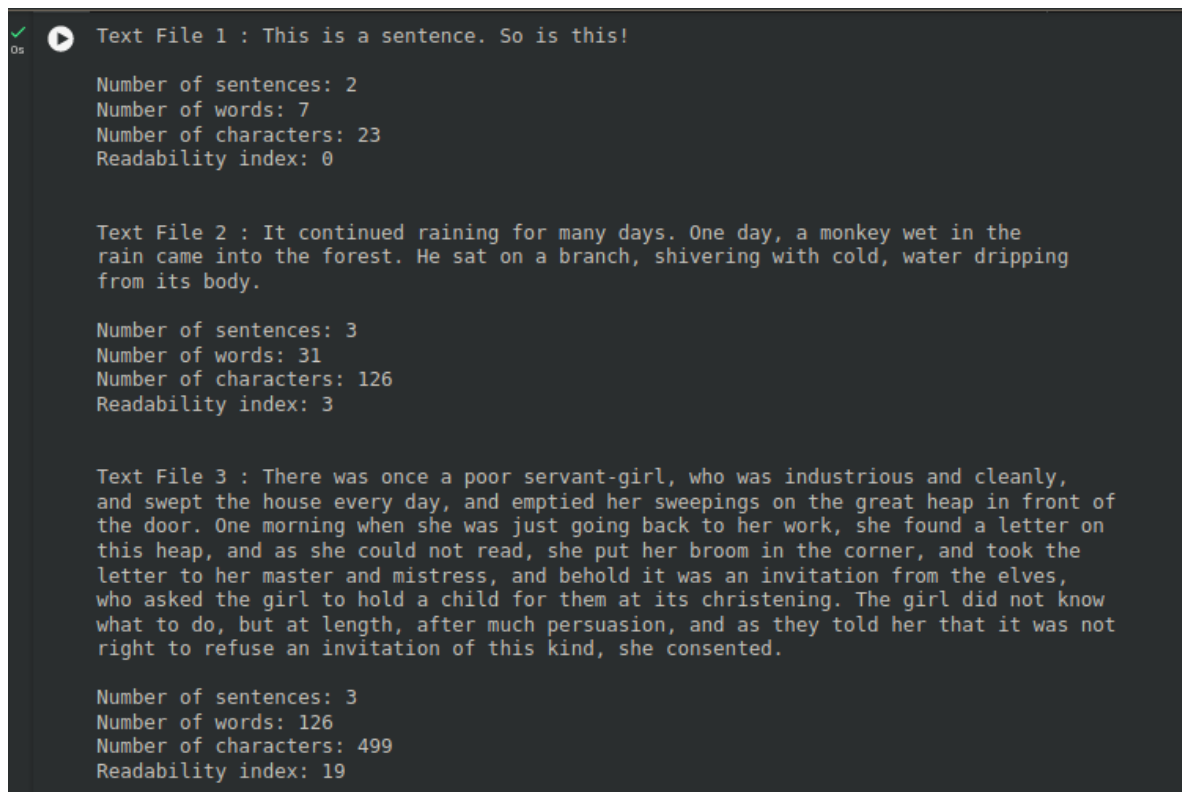
def return_filepath():
    base_dir = '/content/drive/MyDrive/AI/'
    # Get a list of all .txt files in the directory
    text_files = [os.path.join(base_dir, f) for f in os.listdir(base_dir) if f.endswith('.txt')]
    return text_files

def main():
    text_files = return_filepath()
    for text_file in text_files:
        with open(text_file, 'r') as file:
            text = file.read()
            sentences = count_sentences(text)
            words = count_words(text)
            characters = count_characters(text)
            ari = calculate_ari(text)
            print_result(text_file, text, sentences, words, characters, ari)

```

3 Output

Here is the output of the Python script:



```

Text File 1 : This is a sentence. So is this!

Number of sentences: 2
Number of words: 7
Number of characters: 23
Readability index: 0

Text File 2 : It continued raining for many days. One day, a monkey wet in the
rain came into the forest. He sat on a branch, shivering with cold, water dripping
from its body.

Number of sentences: 3
Number of words: 31
Number of characters: 126
Readability index: 3

Text File 3 : There was once a poor servant-girl, who was industrious and cleanly,
and swept the house every day, and emptied her sweepings on the great heap in front of
the door. One morning when she was just going back to her work, she found a letter on
this heap, and as she could not read, she put her broom in the corner, and took the
letter to her master and mistress, and behold it was an invitation from the elves,
who asked the girl to hold a child for them at its christening. The girl did not know
what to do, but at length, after much persuasion, and as they told her that it was not
right to refuse an invitation of this kind, she consented.

Number of sentences: 3
Number of words: 126
Number of characters: 499
Readability index: 19

```

Figure 1: Output of the Python script