CIS 41A Lab 6: Dictionaries and data structures

Write a program that compares sample text files in several Western languages to analyze the frequency of vowels in these languages.

The program reads in 5 text files which are Wikipedia pages on Pablo Picasso in English, French, German, Italian, and Spanish. The program prints out the number of A, E, I, O, U letters (case insensitive) that are in each file, and prints the percentage of vowels used in each file.

The input text files are in lab6input.zip.

Follow these steps to complete the assignments:

1. Extract the zip file into 5 text files

2. Write a main function that:

- has a list of the 5 filenames

- in a loop, calls the function in step 3 and uses the list of filenames to pass to the function one filename at a time

- calls the function in step 4 to print the result

3. Write an analyze function that reads all characters of one file and keeps count of each vowel.

4. Write a function that prints:

- the count of each vowel in each language

- the ratio of (total vowels / total letters), as a percentage

It is up to you determine how data is passed from one function to another. Choose the simplest way and don't use any global variable.

Because some characters in the files are non-ascii characters, you need to tell Python to work around these characters by providing a second argument to the file open function: open (filename, encoding = "latin-1")

Here is the partial program output for you to check your work, some fields are left blank so you can experience the excitement of finding the answers:

English

A: 2870

E: 3354

I: 2750

O: 2337

U: 835

Total vowels / total letters: 38.72%

French

A: 2767

E: 4574

I: 2336

O: 1651

U: 1668

Total vowels / total letters: 40.92%

German

A:

E:

I: 6700

O:

U:

Total vowels / total letters:

Italian

A: 2845

E:

I:

O:

U:

Total vowels / total letters:

Spanish

A:

E:

I:

O: 7310

U:

Total vowels / total letters: