09.4 - File Analysis

Write a program that processes the two provided text files (`python_1.txt` and `python_2.txt`) and produces the following output files.

1. `python_1_word_frequency.txt` This file should contain all of the unique words in the file `python_1.txt` printed one per line in alphabetical order. Each word should be followed by a colon, a space, and the number of times that word appears in `python_1.txt`.

2. `python_2_word_frequency.txt` This file should contain all of the unique words in the file `python_2.txt` printed one per line in alphabetical order. Each word should be followed by a colon, a space, and the number of times that word appears in `python_2.txt`.

3. `common_words.txt` This output file should contain all of the words that appear in both `python_1.txt` and `python_2.txt`, printed one per line in alphabetical order.

4. `eitherbutnotboth.txt` This file should contain all of the words that appear in either `python_1.txt` or `python_2.txt` but no words that appear in both. The words should be printed one per line in alphabetical order.

For comparison purposes, all the words should be converted to lower case and all leading and trailing punctuation should be removed. This first five lines in each of the output files is shown in Figure 1. Save your program as `file_analysis_login.py`, where `login` is your Purdue login. Then submit it along with a screenshot showing the first few lines of **all 4** output files.

Editor - python_1_word_frequency.txt

```
a: 5
able: 1
about: 1
additional: 1
after: 1
```

(a)

Editor - python_2_word_frequency.txt

```
1: 1
a: 21
allows: 1
also: 1
an: 1
```

(b)

Editor - common_words.txt

```
a
also
an
and
application
```

(c)

Editor - eitherbutnotboth.txt

```
1
able
about
additional
after
```

(d)

Figure 1: The first five lines of output files (a) `python_1_word_frequency.txt`, (b) `python_2_word_frequency.txt`, (c) `common_words.txt`, and (d) `eitherbutnotboth.txt`.

Hints:

- The `string` module includes a `punctuation` string which contains all of the ASCII punctuation characters.
- It might be helpful to write a function that accepts a filename as an argument and returns a list of all the words in that file. Then have another function that accepts a list of words as an argument and returns a dictionary with the unique words in the list as keys and the number of times each word appears in the file as values.
- You can use the `sorted()` function to get a sorted list of the keys and values in a dictionary.
- You can use the `set()` function to create a set from the keys in a dictionary.