School of Information Systems
Academic Year 2019/20
Term I

IS111: Introduction to Programming

Lab Exercise 11: File Input/Output

Optional submission deadline: **Sun 10 Nov 2019 @ 2359 hrs**

INSTRUCTIONS

You are expected to read through the accompanying lesson 11 notes before attempting lab exercise 11. You should attempt the questions using **Visual Studio Code**. Each question should be submitted on a different file.

Do not submit Level 1 questions; use the discussion forum for that.

To submit, you should:

1. Name your 3 files as **week11-qn2-1.py**, **week11-qn2-2.py** and **week11-qn3-1.py**.
2. Submit the 3 .py files to **eLearn → Assignments** by the stipulated deadline.

Guidelines for Forum Discussion

You can refer to this link for guidelines on how to ask questions in a discussion forum. For instance, if someone has posted a solution that is exactly the same as yours (**or very similar to yours**), then you need not post your solution again. You are highly encouraged to help your coursemates with their questions, so that everyone can learn together.

Level 1 Questions (no submission required; use the discussion forum)

1.1 Count Number of Words

Write a Python function `count_words` that takes in the following parameter:

- `filename` (type: `str`): This is a non-empty string that contains the filename of a text file.

The function should return the total number of words in the given file. You may assume that each word in the file is separated by a space.

You can use the following code snippet to test your code:

```
count_words('lab-11-shakespeare.txt') # returns 211
```

1.2 All the World is a Stage

Write a Python function `print_file` takes in the following three parameters:

- `filename` (type: `str`): This is a non-empty string that contains the filename of a text file.
- *n* (type: `int`): This is a positive integer.
- *m* (type: `int`): This is a positive integer such that *m* `>=` *n*.

The function should print out every n^{th} line of the file, up to (and including) the first *m* lines of the file. You may assume that the first line of the file is considered *n*=1.

Here is how the output will look like when `print_file('lab-11-shakespeare.txt', 3, 10)` is called.

```
They have their exits and their entrances,
Mewling and puking in the nurse's arms.
Unwillingly to school. And then the lover,
```

Level 2 Questions

2.1 Internet Access at Home

The file `lab-11-internet.csv` is downloaded from data.gov.sg, and contains information about the types of Internet connections used at home, between 2008 to 2015.

The first few lines of the csv file are as follows:

```
year,type_of_connection,internet_access
2008,fixed broadband,91
2008,wireless broadband,8
2008,fibre broadband,na
2009,fixed broadband,92
```

Write one or multiple Python programs to answer the following questions about the data:

1. What are the different types of Internet connections that are available?
2. In which year was fibre broadband introduced?
3. In which year did the fixed broadband percentage fall below the wireless broadband for the first time?

The expected answers to the above questions are as follows:

1. fixed broadband, wireless broadband, and fibre broadband
2. 2012
3. 2013

2.2 Mobile Data Usage

The file `lab-11-mobile.csv` is downloaded from data.gov.sg, and contains information about the mobile data usage between 2004 and 2019.

The first few lines of the csv file are as follows:

```
quarter,volume_of_mobile_data
2004-Q3,0.000384
2004-Q4,0.000543
2005-Q1,0.00062
2005-Q2,0.000634
```

Note that the unit of measure of the volume of data is in petabytes.

Write a program to process the given csv file, and aggregate the volume of mobile data by the year (i.e., adding all the quarters together). Save your output to a csv file `lab-11-mobile-output.csv`, in the following format, whereby the years are in ascending order and the volume of mobile data is rounded off to 5 decimal places:

```
year,volume_of_mobile_data
2004,0.00093
2005,0.00277
2006,0.00714
```

```
2007,0.19714
2008,1.54372
...
```

For years where you may have incomplete data (e.g., year 2004 and year 2019), you may sum up the partial data and provide it in the output file.

3 Brownie Points 🍪

3.1 Secondary School Enrolment

The file `lab-11-secondary.csv` is downloaded from data.gov.sg, and contains data about the secondary school enrolment by level and by course.

The first few lines of the csv file are as follows:

```
year,level,course,sex,enrolment_secondary
1980,Secondary 1,Express,MF,45489
1980,Secondary 1,Express,F,22509
1980,Secondary 1,Normal (Acad),MF,0
1980,Secondary 1,Normal (Acad),F,0
```

Write a program to process the given csv file, and aggregate the total number of **male** students (i.e., the difference in numbers between 'MF' and 'F' in the sex column) who are enrolled in a secondary school (of any level), by the year. Save your output to an output file `lab-11-secondary-output.csv`, in the following format:

```
year,num_male_students
1980,76765
1981,78209
1982,77813
...
```

💾 25 Oct 2019