Assignment 8 – Web Scraping Course Information

Goals

- Practice web scraping by using the Beautiful Soup module
- More practice writing to text files

Requirements

- Using http://classes.usc.edu/term-20181/classes/itp/ (or any other program from the USC course catalog website - http://classes.usc.edu/term-20181/)
- Write information of classes based off the number of units specified by a user to a text file.
    - You may create any functions necessary to implement this program, however a **`main`** function is required (i.e. no global constants or code, or code outside of functions aside from a call to main)
    - Ask the user for how many units they'd like to find classes for.
        - Error check that the user entered a valid number of units (valid range depends on which program you select)
    - Use the **`BeautifulSoup`** module to parse the webpage and write each class that contains the user specified number of units
        - For every found class, also write information about each section. Include each section's time, number of registered students, and instructor name.
    - Write the information to the result file in an organized, clear format.
- Hints:
    - Each course has all its information grouped into its own **`div`** tag. How can you find all these **`div`** tags?
    - Each course's section information is display in a series of table row **`tr`**.
    - If you want to get the inner most text in a series of nested tags, use **`.text`**
      For example, if the variable **`tag`** contains the following:
      `<div><strong>Fight on!</strong></div>`
      then **`print(tag.text)`** will display the following
      `Fight on!`
    - You can make HTML source code easier to read format by copying all the HTML code and pasting it into this website: http://htmlformatter.com/
    - For more information, you can consult the documentation for Beautiful Soup: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

Sample Output

Main Program:

```
Enter the number of units you wish to search for classes by (1-4): 17
*Invalid input, please try again.
Enter the number of units you wish to search for classes by (1-4): 3
See 'results.txt' for your results.
```

results.txt:

```
Here are all of the ITP classes that are 3.0 units:

ITP 300: Database Web Development (3.0 units)
12:30-1:50pm, 0 of 33, Dung Nguyen

ITP 305: Advanced 3D Modeling, Animation, and Special Effects (3.0 units)
12:00-1:50pm, 0 of 16, Lance Winkel

ITP 308: Computer-Aided Design for Bio-Mechanical Systems (3.0 units)
5:00-7:50pm, 0 of 33, Raymond Kim

ITP 310: Design for User Experience (3.0 units)
11:00-1:50pm, 0 of 26, Blessing Yen
9:30-11:50am, 1 of 45,

ITP 325: Ethical Hacking and Systems Defense (3.0 units)
5:00-7:50pm, 0 of 27, Jennifer Kassar

ITP 341: App Development for Phones and Tablets (3.0 units)
12:00-1:50pm, 1 of 38, Robert Parke

ITP 342: Mobile Application Development (3.0 units)
12:00-1:50pm, 0 of 60, Trina Gregory

ITP 344: Advanced Topics in Mobile App Development (3.0 units)
6:00-8:50pm, 0 of 17, Spartak Buniatyan

ITP 365: Managing Data in C++ (3.0 units)
2:00-3:50pm, 0 of 57, Nathan Greenfield
10:00-11:50am, 0 of 40, Nathan Greenfield
```

ITP 368: Programming Graphical User Interfaces (3.0 units)
12:00-1:50pm, 0 of 28, Kendra Walther
12:00-1:50pm, 0 of 40, Kendra Walther

ITP 370: Information Security Management (3.0 units)
6:00-7:20pm, 0 of 22, Mike Cassar

ITP 375: Digital Forensics (3.0 units)
9:00-10:20am, 1 of 27, Jennifer Kassar

ITP 382: Mobile Game Programming (3.0 units)
7:00-9:50pm, 1 of 22, Michael Sheehan

ITP 405: Professional Applications and Frameworks in Web Development (3.0 units)
6:00-8:50pm, 0 of 36, David Tang

ITP 411: Multimedia and Video Production (3.0 units)
9:00-11:50am, 1 of 25, Larry Jordan

ITP 422L: Configuring Enterprise Resource Planning Systems (3.0 units)
5:00-7:50pm, 0 of 26, Bhargav Oza

ITP 435: Professional C++ (3.0 units)
2:00-3:20pm, 1 of 45, Sanjay Madhav
2:00-3:20pm, 1 of 60, Sanjay Madhav

ITP 439: Compiler Development (3.0 units)
5:00-6:20pm, 0 of 35, Sanjay Madhav

ITP 447: Mobile Device Security and Forensics (3.0 units)
6:00-8:50pm, 0 of 16, Pierson Clair

ITP 454x: Enterprise Resource Planning, Design, and Implementation (3.0 units)
5:00-7:50pm, 0 of 28, Richard Vawter

ITP 479: Cyber Law and Privacy (3.0 units)
5:30-8:20pm, 0 of 28, Benyomin Forer

ITP 487: Data Warehouses (3.0 units)
2:00-4:50pm, 0 of 37, Mike Lee

ITP 489: In-Memory Database Systems for Real Time Analytics (3.0 units)
2:00-3:20pm, 0 of 33, Richard Vawter

Deliverables and Submission Instructions

- Create a folder on your computer called
  ITP115_a#_lastname_firstname
  (replace # with this lab number)
- Inside the folder, include your python source code
- Compress the folder (make a zip file) called
  ITP115_a#_lastname_firstname.zip
  (replace # with this assignment number)
- Upload zip file to Blackboard site for our course

Grading

Item	Points
Configuring BeautifulSoup and downloading page	2
Unit user input and error checking	3
Writing all courses to a file	5
Writing all sections to a file	5
Total*	**15**

** Points will be deducted for poor code style, or improper submission.*