UNIVERSITÉ Concordia UNIVERSITY

ASSIGNMENT 2
Assignment Posted: February 17, 2024
Assignment Due: March 09, 2024, before 11.59pm
Final Deadline with 20% flat Penalty: March 12, 2024, before 11.59pm

"Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom."

— Clifford Stoll

Description:

In this assignment, you will venture into the practical realm of Python programming. While a comprehensive mastery of Python is not a prerequisite, this task offers you an invaluable opportunity to delve into various aspects of Python covered in the class. Beyond reinforcing familiar concepts, you'll also encounter a few new elements and concepts, enriching your hands-on experience with the language and expanding your programming skill set.

NOTE: Please note that the assignment document may seem lengthy, but the description of the assignment itself is concise and clear. The document primarily consists of valuable hints and advice designed to streamline development of your Python program. Additionally, sample code snippets are included to facilitate utilization of libraries, making the coding process more accessible.

In this assignment, you will dive into practical application of Python programming to create a web scraping project. Web scraping, a technique used to extract valuable information from websites, is a key skill in the realm of data collection and analysis. Much like how major platforms leverage data for personalized recommendations, this assignment provides you with the opportunity to explore Python's capabilities in harvesting information from the web. While the complexity of this project may not rival large-scale systems, it serves as a hands-on exercise in web scraping, enhancing your proficiency in Python programming.

Ethical web scraping involves conducting data extraction from websites in a responsible and legal manner, respecting the terms of service of the targeted sites. In this educational example, we emphasize the importance of adhering to ethical standards while performing web scraping activities. The purpose of this assignment is solely educational, aimed at providing students with practical experience in web scraping techniques and data analysis. It is essential to highlight that the code and techniques demonstrated in this assignment are intended for educational purposes, and students are encouraged to apply these skills responsibly and ethically in compliance with relevant laws and regulations. Understanding and practicing ethical web scraping is crucial for fostering a respectful and responsible approach to data extraction. We will be working with ***refdlagdeals.com*** as a case study.

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5):
```

Figure 1. Main menu

Your application must display the welcome message and a main menu as depicted in Fig.1 below. The application will have five options and you may assume a perfect user who will always enter an appropriate type of input, for instance an integer for number input.

In this assignment, we leverage two powerful libraries for efficient web scraping in Python: **requests** and **BeautifulSoup**. The _requests_ library is employed to send HTTP requests and retrieve HTML content from the specified website. _BeautifulSoup_ facilitates the parsing and extraction of relevant information from the HTML, simplifying the process of navigating through the webpage's structure.

To aid your understanding and implementation, a sample code snippet is thoughtfully provided with this assignment. This code serves as a foundational template, demonstrating how to utilize these libraries effectively for data scraping. By exploring and customizing this sample code, you'll gain practical insights into the mechanics of web scraping and be better equipped to tackle the assignment with confidence.

Implementation Specifications:

Detailed behavior of about each menu option is mentioned below with screenshots.

1. **Display Latest Deals**

   Selecting this option allows a user to view the most recent deals scraped from the designated website. The program will present a neatly formatted display, showcasing essential details such as the store name, item description, voting counts, username, timestamp, category, replies, views, and URL link for each deal. This feature provides a quick snap of the latest and most relevant deals gathered during the web scraping process. It serves as a convenient way to stay updated on the current offerings without navigating through the entire dataset. The total deals found should be printed on the first line followed by details on each deal. A sample of the format is given below.

```
    Store: EQ Audio
     Item: Focal Aria 926 3-way floor standing speakers $1249 each (reg $3099)
    Votes: +12
 Username: DavidY
Timestamp: Feb 17th, 2024 9:27 am
 Category: Computers & Electronics
  Replies: 8
    Views: 1,305
      URL: https://forums.redflagdeals.com//eq-audio-focal-aria-926-3-way-floor
```

Figure 2 (a). A sample deal output

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 1

Total deals found: 55

    Store: EQ Audio
     Item: Focal Aria 926 3-way floor standing speakers $1249 each (reg $3099)
    Votes: +12
 Username: DavidY
Timestamp: Feb 17th, 2024 9:27 am
 Category: Computers & Electronics
  Replies: 8
    Views: 1,305
      URL: https://forums.redflagdeals.com//eq-audio-focal-aria-926-3-way-floor-standing-speakers
-----------------------------------------
    Store: Freedom Mobile
     Item: Prepaid - $119 for 12 months wireless service with nationwide calling and 15GB data
    Votes: +206
 Username: User090179
Timestamp: Dec 12th, 2023 8:31 am
```

Figure 2(b). Display Latest Deals

2. Analyze Deals by Category

Selecting this option will present the user with a comprehensive list of all available categories, along with the corresponding number of deals within each category. This insightful overview allows you to quickly assess the distribution of deals across different segments. By displaying the quantity of deals per category, this option offers a convenient snapshot of the dataset's categorical landscape. It serves as a valuable tool for users seeking a broad understanding of variety and volume of deals present in the scraped data.

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 2

Deals by Category:

Computers & Electronics: 29 deals
              Groceries:  8 deals
          Home & Garden:  2 deals
     Financial Services:  3 deals
      Beauty & Wellness:  4 deals
             Automotive:  4 deals
          Entertainment:  4 deals
          Kids & Babies:  1 deals
```

Figure 3. Analyze Deals by Category

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 3
Enter the number of top stores to display: 5

Top Stores:

        No Frills:  6 deals
    Freedom Mobile:  4 deals
        Amazon.ca:  4 deals
         Best Buy:  4 deals
     Lenovo Canada:  3 deals
-------------------------------------------------------


***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5):
```

Figure 4. Find Top Stores

3. Find Top Stores

Upon selecting this option, the program prompts the user to input the desired number of top stores to display. Once the input is provided, the system analyzes the dataset, identifying and presenting top stores based on the number of deals associated with each store. This interactive feature allows you to customize your exploration, focusing on the stores that have the highest volume of deals. It serves as an effective tool for discerning popular and prolific sources of deals within the scraped data, contributing to a more personalized and targeted analysis.

4. Log Deal Information

When this option is selected, the program first displays all available categories. A user is then prompted to enter the number matching their chosen category. Subsequently, the program identifies and logs all the deal links within the selected category to a file named 'log.txt'. This practical feature streamlines the process of documenting and organizing deal information based on the specific categories. By offering a targeted approach to log deals, this option provides a convenient means of curating and storing information tailored to user-defined interests.

NOTE: A fundamental strength of Python lies in its commitment to readability, a core principle encapsulated in the Zen of Python. Every output produced by your program must be formatted neatly. 'Readability counts.' You will lose points if any part of your code or output produced by it is not neatly formatted.

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 4

List of Categories:

 1. Computers & Electronics
 2. Groceries
 3. Home & Garden
 4. Financial Services
 5. Beauty & Wellness
 6. Automotive
 7. Entertainment
 8. Kids & Babies
Enter the number corresponding to the category: 4
All the links have been logged successfully.

***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5):
```

Figure 5. Log Deal Information

5. Exit

The application must exit gracefully, and all the resources must be released upon pressing the five ('5') key.

Submission:

Assignment must be submitted only through Moodle. No other form of submission will be considered. Please create a zip file containing your Python code and a readme file (txt). The zip file should be named Assignment#2_YourStudentID. In the readme file document, the features and functionality of the application, and anything else you want grader to know.

Additional Information

❖ You can use the code provided during the tutorial or lecture sessions to get started.

Extra Credit (10% Points)

A more elaborate application (Deal Alerts based on specific types of deals/criteria, receiving notifications when new matching deals are scraped., Deal filtering as well as deal sorting, *etc.*) at grader's discretion.

Evaluation Procedure

You MUST demonstrate your program to the grader during the scheduled demo time. You must run your submitted code, demonstrate its full functionality and answer questions about the programming aspects of your solution. Major marking is done on the spot during demonstration. Your code will be further checked for structure, non-plagiarism, *etc.*

However, ONLY the demonstrated submissions will receive marks. Other submissions will not be graded. **THIS ASSIGNMENT MUST BE SUBMITTED INDIVIDUALLY.**

HINTS:

1. Use a dict to collect all the deals for better organization and access.
2. Implement separate classes to demonstrate the object-oriented principles in Python.
3. Include error handling for file operations (e.g., opening, reading, and writing files).
4. Implement various modules and perform apt imports for code reusability.

PITFALLS:

1. Be mindful of file handling errors; check for file existence before reading.
2. Avoid hardcoded values where possible; use constants for better maintainability.
3. Scraping too frequently can strain your resources and lead to poor performance.

IMPORTANT!!!

Markers will receive a straightforward spreadsheet containing the specified criteria outlined earlier. There are no undisclosed or hidden requirements. It is emphasized that having a functional version of a somewhat limited program is preferable to having a non-functional version attempting to cover every aspect. Therefore, prioritize ensuring that your code can be successfully compiled and executed. Evaluating an assignment that doesn't run at all is virtually impossible.

Sample Run

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 1

Total deals found: 55

    Store: EQ Audio
     Item: Focal Aria 926 3-way floor standing speakers $1249 each (reg $3099)
    Votes: +12
 Username: DavidY
Timestamp: Feb 17th, 2024 9:27 am
 Category: Computers & Electronics
  Replies: 8
    Views: 1,305
      URL: https://forums.redflagdeals.com//eq-audio-focal-aria-926-3-way-floor-standing-
           speakers-1249-each-reg-3099-2676450/
```

```
-----------------------------------------
    Store: Freedom Mobile
     Item: Prepaid - $119 for 12 months wireless service with nationwide calling and 15GB data
    Votes: +206
 Username: User090179
Timestamp: Dec 12th, 2023 8:31 am
 Category: Computers & Electronics
  Replies: 1,110
    Views: 169,887
      URL: https://forums.redflagdeals.com//freedom-mobile-prepaid-119-12-months-wireless-
           service-nationwide-calling-15gb-data-2661833/
-----------------------------------------
    Store: No Frills
     Item: Swanson Oven Easy Meals or Family Skillets $1.88 (reg. $11.49) with printed coupon
    Votes: +56
 Username: hill15
Timestamp: Feb 13th, 2024 3:25 pm
 Category: Groceries
  Replies: 65
    Views: 10,817
      URL: https://forums.redflagdeals.com//no-frills-swanson-oven-easy-meals-family-skillets/
-----------------------------------------
.
.
.
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 2

Deals by Category:

Computers & Electronics: 29 deals
             Groceries:  8 deals
          Home & Garden:  2 deals
     Financial Services:  3 deals
      Beauty & Wellness:  4 deals
             Automotive:  4 deals
          Entertainment:  4 deals
          Kids & Babies:  1 deals
------------------------------------------------------

***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 3
Enter the number of top stores to display: 25

Top Stores:
        No Frills:  6 deals
   Freedom Mobile:  4 deals
        Amazon.ca:  4 deals
         Best Buy:  4 deals
```

```
            Lenovo Canada:  3 deals
                EQ Audio:  2 deals
               Shopper+:  2 deals
            Amazon.com:  2 deals
                Walmart:  2 deals
                OnePlus:  2 deals
              Air Miles:  2 deals
       PlayStation Store:  2 deals
                 Costco:  2 deals
          Princess Auto:  2 deals
             Distributel:  2 deals
          Public Mobile:  2 deals
                Samsung:  2 deals
                 adidas:  2 deals
                  Shell:  2 deals
        B&H Photo Video:  1 deals
               Uppababy:  1 deals
            Wealthsimple:  1 deals
              Formula 1:  1 deals
                   Fizz:  1 deals
                TextNow:  1 deals
--------------------------------------------------------
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 2

Deals by Category:

Computers & Electronics: 29 deals
              Groceries:  8 deals
           Home & Garden:  2 deals
      Financial Services:  3 deals
       Beauty & Wellness:  4 deals
              Automotive:  4 deals
           Entertainment:  4 deals
           Kids & Babies:  1 deals
--------------------------------------------------------
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 3
Enter the number of top stores to display: 5

Top Stores:

             No Frills:  6 deals
         Freedom Mobile:  4 deals
               Amazon.ca:  4 deals
                Best Buy:  4 deals
           Lenovo Canada:  3 deals
--------------------------------------------------------
```

```
***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 4

List of Categories:

 1. Computers & Electronics
 2. Groceries
 3. Home & Garden
 4. Financial Services
 5. Beauty & Wellness
 6. Automotive
 7. Entertainment
 8. Kids & Babies
Enter the number corresponding to the category: 4
All the links have been logged successfully.

***** Web Scraping Adventure *****
1. Display Latest Deals
2. Analyze Deals by Category
3. Find Top Stores
4. Log Deal Information
5. Exit
Enter your choice (1-5): 5
Exiting the program. Goodbye!
```

REMINDER:

As you work on your web scraping project, we highly encourage you to comment your code thoroughly. Comments are an essential part of good programming practice, and they serve multiple purposes including documentation, clarity, troubleshooting, etc. Please _use inline comments to explain complex logic or non-trivial sections of your code_. Provide _docstrings for functions and methods, explaining their parameters, return values, as well as functionality_. Comment on important decisions, assumptions, or limitations in your code.

Grading Rubric for Assignment 2

Total	100 pts
Display Latest Deals	_15 pts_
Analyze Deals by Category	_10 pts_
Find Top Stores	_10 pts_
Log Deal Information	_15 pts_
Formatted Output and Fully Commented Code	_20 pts_
Demo QnA by the Grader	_30 pts_