Assignment: Web Scraping with Python BeautifulSoup and LinkedIn

Objective:

The objective of this assignment is to demonstrate your skills in web scraping using Python and BeautifulSoup library. You will be scraping data from LinkedIn, a popular professional networking platform, to extract specific information related to user profiles.

Instructions:

1. Set up the environment:

- Install Python if you haven't already.
- Install the BeautifulSoup library by running the command: `pip install beautifulsoup4`.

2. Scraping LinkedIn Profiles:

Your task is to scrape information from multiple LinkedIn profiles based on specific criteria. Follow the steps below:

a. Target Profiles:

- Select a specific industry or domain of interest (e.g., technology, marketing, finance, etc.).
- Choose at least five LinkedIn profiles that belong to professionals within that industry.

b. Data to Scrape:

- For each profile, extract the following information:
  - Name
  - Headline/Title
  - Location
  - Current Company

- Education
- Summary/About
- Skills

c. Implementation:

- Write a Python script that performs the web scraping using BeautifulSoup.
- Use appropriate HTML tags and attributes to locate and extract the desired data from the LinkedIn profiles.
- Make sure to handle cases where the information is not available or structured differently on certain profiles.

d. Output:

- Store the scraped data in a suitable format, such as a CSV file.
- Include headers for each column representing the extracted information.

3. Data Validation:

To ensure the accuracy and completeness of the scraped data, perform the following validations:

a. Data Integrity:

- Check if all the required fields have been successfully scraped for each profile.
- Handle cases where any field is missing or contains invalid data.

b. Data Consistency:

- Verify the consistency of data across profiles.
- Identify and handle any inconsistencies or discrepancies in the scraped data.

4. Additional Enhancements (Optional):

If you would like to challenge yourself further, consider implementing the following enhancements to your web scraping script:

a. Pagination:

- Modify your script to scrape information from multiple pages of search results.
- Implement pagination to scrape profiles beyond the first page of search results.

b. Advanced Data Extraction:

- Extract additional information, such as work experience, contact information, or recommendations, from the LinkedIn profiles.
- Use advanced techniques to handle more complex HTML structures or dynamic content.

Submission Guidelines:

- Create a GitHub repository to host your web scraping project.
- Include a README file with clear instructions on how to run your script and any additional notes.
- Upload the link to your GitHub repository or share the repository directly with the evaluator.

Evaluation Criteria:

- Successful extraction of the required information from LinkedIn profiles.
- Accuracy and completeness of the scraped data.
- Proper handling of missing or inconsistent data.
- Implementation of any additional enhancements.
- Code quality, readability, and adherence to best practices.
- Documentation and clarity of instructions.

Note: Remember to respect LinkedIn's terms of service and usage policies while performing web scraping.