Crawling Assignment: LinkedIn Waste Management Requirements Collection

Objective

The objective of this task is to crawl LinkedIn for posts related to waste management using relevant keywords or hashtags. The goal is to extract actionable marketplace requirements from the posts and categorize them into predefined waste management categories.

Task Description

1. **Search Strategy**:
    - Use targeted **keywords** such as "waste management requirements," "waste disposal needed," or hashtags like #PlasticWaste, #EwasteManagement, #SustainableWaste, etc.
    - Focus on extracting **requirements for marketplace listings** from relevant posts.
2. **Data Points to Extract**:
    - **Actual Requirement**: The specific waste management need mentioned in the post.
    - **Requirement Poster Information**: Include name, designation, and company (if available).
    - **Date of Post**: The date when the post was made.
    - **Category of Requirement**: Classify the requirement into one of the following categories:
        - **Plastic Waste (A)**
        - **E-waste (B)**
        - **Bio-medical Waste (C)**
        - **Construction and Demolition Waste (D)**
        - **Battery Waste (E)**
        - **Radioactive Waste (F)**
        - **Other Hazardous Waste (G)** (Specify type, if mentioned) for example –
            1. **Chemical Waste**: Residues from paints, solvents, or industrial chemicals.
            2. **Pesticides and Herbicides**: Expired or unused agricultural chemicals.
            3. **Asbestos**: Materials containing asbestos fibers.
            4. **Industrial Sludge**: Hazardous byproducts from manufacturing processes.
            5. **Contaminated Soil**: Soil polluted with hazardous substances like oil or heavy metals.
        - **Other Non-Hazardous Waste (H)** (Specify type, if mentioned) for example –
            1. **Food Waste**: Organic waste from food production or consumption.

2. **Paper and Cardboard**: Recyclable office or packaging waste.
3. **Textile Waste**: Used fabrics, clothes, or industrial fabric remnants.
4. **Glass Waste**: Broken or discarded glass bottles and products.
5. **Wood Waste**: Sawdust, wooden pallets, or construction leftovers.
6. **Rubber Waste**: Non-hazardous rubber from tires or other sources

3. **Crawling Steps**:
    - Log in to LinkedIn using an authorized account.
    - Utilize LinkedIn's search functionality or API (if access is granted) to query posts using keywords/hashtags.
    - Parse post data to identify and extract the mentioned data points.
4. **Output Format**:
   The output should be structured and saved in a format such as **CSV/JSON**, with the following columns/fields:
    - Post Content
    - Requirement Category (A-H, specify if applicable)
    - Poster Name
    - Poster Designation
    - Poster Company
    - Post Date

Expected Deliverables

- A **data file** (CSV/JSON) containing the structured information.
- Code used for crawling, including comments for readability and reproducibility.