Advanced Programming — Python—

Part two: Advanced Python

HOMEWORK

Instructions for the Homework:

Overview:

In this assignment, you will use Python's re module to extract and process specific information from a more complex text. This task will involve extracting and validating data such as email addresses, phone numbers, dates, and social media posts with regex. You will need to handle a variety of formats and patterns, making the task more challenging.

Tasks:

Task 1: Extract and Normalize Emails

1. **Email Extraction**: Write a function ***extract_emails(text)*** that extracts all email addresses from the text. The email addresses may appear in different formats such as:
    - user[at]domain.com (common obfuscation)
    - user@domain.co.uk
    - Emails with special characters in the domain like user@[web].com

Deliverables:

- A list of extracted email addresses.
- Convert all obfuscated emails ([at] to @) before adding them to the list.

Task 2: Extract and Normalize Phone Numbers

1. **Phone Number Extraction**: Write a function ***extract_phone_numbers(text)*** that extracts all phone numbers from the text. Phone numbers can be in various formats, such as:
    - US format: (555) 444-7890
    - International formats: +1-800-555-2468, +44 20 7946 0958
    - Numbers with periods or mixed characters: 1.800.400.5123, 001 (345) 567-8932

Deliverables:

- A list of extracted phone numbers.
- Normalize all phone numbers to the format: +<country code> <area code> <number>, e.g., +1 800 5552468.

Task 3: Extract and Transform Dates

1. **Date Extraction**: Write a function ***extract_dates(text)*** to extract all dates. Dates may appear in different formats, such as:
    - 15/08/1980
    - 10.12.2024
    - 2019/11/29
    - Friday, the 31st of January, 2025
    - April 23rd, 2015
2. **Date Transformation**: Write a ***function transform_dates(date)*** that converts the extracted dates into the standard YYYY-MM-DD format.

Deliverables:

- A list of extracted dates in the original format.
- A list of the same dates transformed into the YYYY-MM-DD format.

Task 4: Password Validation

1. **Password Validation**: Write a function ***validate_password(password)*** to validate passwords. The passwords in this text must meet the following criteria:
    - At least 10 characters long.
    - Contains at least one uppercase letter.
    - Contains at least one lowercase letter.
    - Contains at least one digit.
    - Contains at least one special character (!@#$%^&*).

Deliverables:

- Validate the passwords Secure!2021 and Research@1234 from the text.
- Return True for valid passwords and False for invalid ones.

Task 5: Extract Hashtags from Social Media Posts

1. **Hashtag Extraction**: Write a function ***extract_hashtags(text)*** to extract all hashtags (words starting with #) from the social media posts included in the text.

Deliverables:

- A list of all extracted hashtags.

Bonus Task: Extract Important Dates

1. **Bonus**: Write a function ***find_upcoming_events(text, year)*** that finds all event-related dates (e.g., meetings, deadlines, tasks) that are scheduled to occur in the specified year. For example, if the year is 2024, extract dates and events related to that year.

Deliverables:

- A list of events occurring in the given year, along with their respective dates in YYYY-MM-DD format.

Submission Instructions:

- Submit a Python script (.py file) containing your solutions.
- Each function should be clearly documented with comments explaining your approach.
- Ensure your script outputs the results for each task when executed (emails, phone numbers, transformed dates, validated passwords, extracted hashtags, etc.).