Université Libanaise
Faculté d'Information
Br. I

الجامعة اللبنانية
كلية الاعلام
الفرع الاول

HB

امتحانات: الفصل الرابع
من العام الجامعي 2026/2025

المرحلة: الاجازة	الدورة الاولى	المادة: Data Scraping
السنة المنهجية: الثانية		المدة: ساعتان
الاختصاص: علم البيانات		اسم الاستاذ: د. محمد أبوطعام

Part I: Choose the correct answer (20 pts)

1. What does the term web scraping usually refer to?
   a) Browsing the web for information
   b) Manually copying and pasting information from websites
   c) Downloading files from the Internet
   d) A process that involves automation to gather information from the Internet

2. Beautiful Soup is used for?
   a) Parsing HTML
   b) Fetching HTML
   c) To extract patterns from text
   d) To execute JavaScript code

3. Which Python library is primarily used for making HTTP requests to fetch a webpage's content?
   a) Beautiful Soup
   b) requests
   c) Pandas
   d) Selenium

4. What does the Beautiful Soup method find() do?
   a) To search through the text
   b) To hide the scraper's IP address
   c) To parse HTML content
   d) To handle JAVASCRIPT rendering

5. What protocol can be used to retrieve web pages using python?
   a) urllib
   b) bs4
   c) http
   d) get

6. What is a python library that can be used to send and receive data over HTTP?
   a) http
   b) urllib
   c) port
   d) header

7. What is a common method to store scraped data?
   a) Printing data to the console
   b) Sending data via email
   c) Storing data in CSV or JSON files
   d) Uploading data to social media

8. What is the purpose of the requests library in web scraping?
   a) To render web pages in a browser.
   b) To handle HTTP requests and responses.
   c) To create visualizations of scraped data.
   d) To automate from submissions

9. What kinds of data can you scrape from the web?
   a) Text data
   b) Images and videos
   c) All type of data
   d) Sentiments and reviews

10. What is the purpose of using regular expressions in web scraping?
    a) To generate HTML code
    b) To scrape data from databases
    c) To extract patterns from text
    d) To render JavaScript code

Part II: Complete the sentences with some of the words from the box below (10 pts)

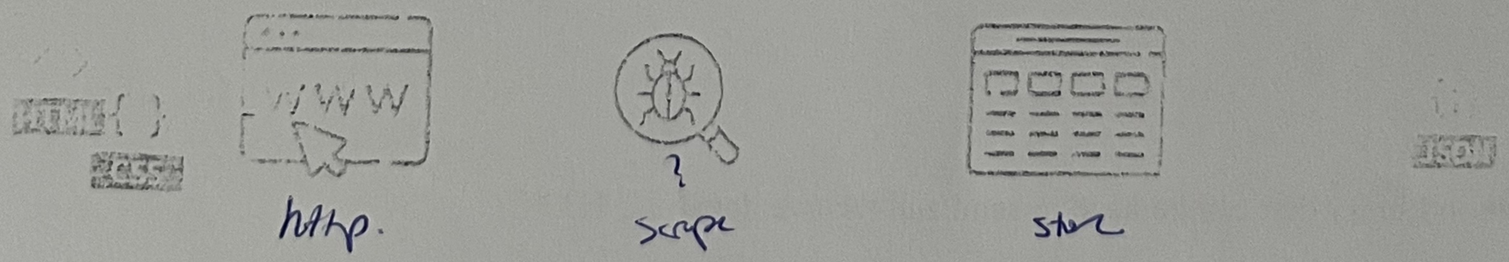
Parsing – Beautiful Soup – Scraping – Inspecting – pip – Crawling

Web scraping is the process of automated information gathering from the Internet.

The Python libraries ---------- and ---------- are essential tools for this task. The web scraping process usually involves three consecutive steps:

1. ---------- the HTML structure of your target site
2. ---------- the data from the Internet
3. ---------- the data to select the information you need

Part III: Explain the process of extracting information from the web in the below figure. (10 pts)



Part IV: Explain in details and show the output of the code: (20 pts)

```
from bs4 import BeautifulSoup
from urllib.request import urlopen
url = "http://olympus.realpython.org/profiles/dionysus"
page = urlopen(url)
html = page.read().decode("utf-8")
soup = BeautifulSoup(html, "html.parser")
print(soup.get_text())
```

Name: Dionysus

Hometown: Mount Olympus

Favorite animal: Leopard

Favorite Color: Wine

Part V: (40 points)

1- What is data scarping? Explain with an example.
2- Why is data scraping useful for data scientist?
3- Is scraping legal? Explain.
4- What is the difference between crawling and scraping?
5- Describe a real-life application where web scraping is used.
6- Describe three python libraries used for web scraping and differ between them.
7- What is the relation between http protocol and web scraping?
8- What is the most preferred programming language for web scraping? Why?