



امتحانات: الفصل الثاني
من العام الجامعي 2024/2023

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

Part I: Choose the correct answer (45 points)

1. 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
2. What is Python IDLE?
 - a) An Integrated Development and Learning Environment
 - b) Perfect tool for a beginning programmer.
 - c) Create and edit Python files and interact with the Python shell
 - d) all of the above
3. 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
4. What is a python library that can be used to send and receive data over HTTP?
 - a) header
 - b) Port
 - c) urllib
 - d) HTTP
5. What is the purpose of Scrapy framework in web scarping?
 - a) Structured and efficient web Scraping
 - b) Creating web forms
 - c) Automating browser actions
 - d) Handling HTTP requests
6. What is the process by which search engines retrieve webpages and build a search index called?
 - a) Scrape
 - b) Parse
 - c) Beautiful Soup
 - d) Spider

7. What does the following block of code do?

- a) Retrieves and displays the webpage
- b) Parses the html content of the "http://www.nytimes.com" webpage
- c) Downloads the webpage

```
url=http://www.nytimes.com
html = urllib.request.urlopen(url).read()
soup = BeautifulSoup(html, 'html.parser')
```

8. What is the Selenium library used for in web scraping?

- a) Scraping static websites
- b) Creating APIs
- c) Automating browser actions
- d) Handling HTTP requests

9. One of the web scraping tools is:

- a) Parsehub
- b) API
- c) Html
- d) HTTP

10. How does the headless browser mode in Selenium work?

- a) It runs a browser with a visible user interface
- b) It runs a browser without a visible user interface
- c) It runs a browser with extra features

11. If the URL doesn't exist, an unsuccessful request might show a status code of:

- a) 200
- b) 404
- c) 100
- d) 500

12. What is the BeautifulSoup library used for in web scraping?

- a) Sending HTTP requests
- b) Rendering JavaScript
- c) Extracting data from HTML and XML

13. What is the rate limiting in the context of web scraping?

- a) Automatically blocking Scraping activities
- b) Sending requests as a controlled rate
- c) Scraping only during specific hours

14. What is web scraping?

- a) Designing web templates
- b) Creating websites
- c) Extracting data from websites
- d) Analyzing website performance

15. What is the purpose of requests library in web scraping?

- a) To create visualizations of scraped data
- b) To render web pages in a browser
- c) To handle HTTP requests and responses

16. Which HTTP method is used for retrieving data in web scraping?

- a) POST

- b) PUT
 - c) GET
 - d) DELETE
17. How can you avoid being blocked while web scraping?
- a) By scraping large amounts of data quickly
 - b) By using multiple IP addresses
 - c) By scraping data from a single website
 - d) By ignoring the robots.txt file
18. 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 files
 - d) Uploading data to social media
19. What is the purpose of Pandas library in web scraping?
- a) Handling HTTP requests
 - b) Rendering JavaScript code
 - c) Extracting data from html and XML
 - d) Manipulating and analyzing data
20. What is the role of re module in web scraping?
- a) Handling HTTP requests
 - b) Parsing JSON responses
 - c) Working with regular expressions
 - d) Creating web forms
21. How can handle websites that load data through JavaScript in web scraping?
- a) By scraping only static content
 - b) By using a headless browser
 - c) By ignoring dynamic content
 - d) By avoiding such websites
22. What protocol can be used to retrieve web pages using python?
- a) Hyper Text Markup Language
 - b) Hyper Tech Markup Language
 - c) Home Tool Markup Language
 - d) Hyperlinks and Text Markup Language

Part II: (55 points)

- 1- What is data scraping?
- 2- Talk about popular data scraping techniques?
- 3- What is web scraping used for (data scraping use cases)?
- 4- What are the legal and ethical considerations of data scraping?
- 5- What are the data scraping tools?
- 6- Based on your web scraper - class project, talk about the limitations that the websites scripting process may face.