

Mock Exam 1

Q1. What is the main purpose of the requests library in web scraping?

- A. Parse HTML
- B. Send HTTP requests
- C. Execute JavaScript
- D. Style web pages

Q2. Which HTTP method is typically used to retrieve a web page?

- A. POST
- B. PUT
- C. GET
- D. DELETE

Q3. What does an HTTP status code 200 indicate?

- A. Redirection
- B. Client error
- C. Server error
- D. Successful request

Q4. Which attribute of a requests response contains the HTML content?

- A. response.content_type
- B. response.html
- C. response.text
- D. response.body

Q5. What is a regular expression mainly used for in scraping?

- A. Styling HTML
- B. Locating patterns in text
- C. Sending HTTP requests
- D. Rendering pages

Q6. Which regex pattern matches one or more digits?

- A. \d
- B. \d*
- C. \d+
- D. [a-z]+

Q7. In HTML, what does the <a> tag represent?

- A. A paragraph
- B. A container
- C. A hyperlink
- D. An image

Q8. Which HTML attribute is commonly used to uniquely identify an element?

- A. class
- B. style
- C. id
- D. name

Q9. What does the DOM represent?

- A. A CSS file
- B. A database
- C. A tree structure of the page
- D. A Python object

Q10. Which BeautifulSoup method returns the **first matching element**?

- A. find_all()
- B. select()
- C. find()
- D. get()

Q11. What does soup.find_all('p') return?

- A. One paragraph
- B. All paragraph tags
- C. Only visible paragraphs
- D. Paragraph attributes

Q12. Which parser is most commonly used with BeautifulSoup?

- A. xml
- B. lxml
- C. html.parser
- D. json

Q13. How do you extract clean text from a BeautifulSoup tag?

- A. .string()
- B. .value
- C. .get_text()
- D. .html()

Q14. What does this CSS selector target: div.product?

- A. All divs
- B. Divs with id "product"
- C. Divs with class "product"
- D. Nested divs

Q15. Which selector finds all links starting with https?

- A. a = https
- B. a[href*="https"]

C. `a[href^="https"]`

D. `a[href$="https"]`

Q16. What is a static website?

A. Uses JavaScript heavily

B. Requires login

C. Content is pre-generated HTML

D. Updates content in real time

Q17. Why should scraping requests be slowed down using `time.sleep()`?

A. Improve parsing speed

B. Avoid server overload

C. Fix HTML errors

D. Increase accuracy

Q18. Which exception is raised by `response.raise_for_status()`?

A. `SyntaxError`

B. `ValueError`

C. `HTTPError`

D. `TypeError`

Q19. What is the role of headers like User-Agent?

A. Encrypt data

B. Simulate a browser

C. Speed up parsing

D. Block requests

Q20. Which tool converts scraped data into tabular format for analysis?

A. NumPy

B. Selenium

C. Pandas

D. Matplotlib

Answer Key

1. B

2. C

3. D

4. C

5. B

6. C

7. C

8. C

9. C

10. C

11. B

12. C

13. C

14. C

15. C

16. C

17. B

18. C

19. B

20. C