
Started on Friday, 28 March 2025, 10:21 AM

State Finished

Completed on Friday, 28 March 2025, 10:29 AM

Time taken 7 mins 45 secs

Marks 13.00/15.00

Grade 86.67 out of 100.00

Question 1

Complete

Mark 1.00 out of 1.00

In Scrapy, what does the start_urls attribute do?

- ☐ a. Stores downloaded pages
- ☒ b. Automatically triggers the parse() method for each URL
- ☐ c. Registers middleware
- ☐ d. Defines URLs for MongoDB

Question 2

Complete

Mark 1.00 out of 1.00

What happens if the next_page selector returns None?

- ☒ a. The crawl ends gracefully
- ☐ b. Scrapy restarts from the first page
- ☐ c. Scrapy skips the URL
- ☐ d. Scrapy throws an error

Question 3

Complete

Mark 1.00 out of 1.00

What is the primary role of the Scrapy Engine?

- ☒ a. Coordinate requests, responses, and items between components
- ☐ b. Render JavaScript-based pages
- ☐ c. Send data to MongoDB
- ☐ d. Manage HTML parsing

Question 4

Complete

Mark 1.00 out of 1.00

What is the purpose of defining an Item class in Scrapy?

- ☐ a. Replaces the parse() method
- ☒ b. Provides structured data containers
- ☐ c. Stores request headers
- ☐ d. Automatically maps to MongoDB

Question 5

Complete

Mark 1.00 out of 1.00

What is the result of calling insert_one() in process_item() but forgetting to close the client in close_spider()?

- ☐ a. MongoDB auto-closes everything safely
- ☐ b. Scrapy crashes
- ☐ c. Data will not be written at all
- ☒ d. It works, but may cause a memory leak or leave open connections

Question 6

Complete

Mark 1.00 out of 1.00

What will happen if you forget to return the item in process_item()?

- ☐ a. MongoDB inserts an empty document
- ☐ b. Scrapy skips it
- ☒ c. Data will not pass to next pipeline or output
- ☐ d. It crashes

Question 7

Complete

Mark 0.00 out of 1.00

Where is the MongoDB connection usually initialized in a Scrapy pipeline?

- ☐ a. parse()
- ☐ b. open_spider()
- ☐ c. __init__()
- ☒ d. settings.py

Question 8

Complete

Mark 1.00 out of 1.00

Which CSS selector correctly extracts quote text?

- ☐ a. `a.tag[href]`
- ☒ b. `span.text::text`
- ☐ c. `div.quote-text::value`
- ☐ d. `div.author`

Question 9

Complete

Mark 1.00 out of 1.00

Which line extracts all tag texts for a quote?

- ☐ a. `q.css('div.tags a.tag').get()`
- ☐ b. `q.select('tags').all()`
- ☒ c. `q.css('div.tags a.tag::text').getall()`
- ☐ d. `q.css('a.tag::href').getall()`

Question 10

Complete

Mark 1.00 out of 1.00

Which method in a Spider is called for each response?

- ☒ a. `parse()`
- ☐ b. `yield()`
- ☐ c. `start_requests()`
- ☐ d. `scrape()`

Question 11

Complete

Mark 1.00 out of 1.00

Which of the following can be returned from a Spider's `parse()` method?

- ☐ a. Only Response objects
- ☐ b. Only Items
- ☒ c. Either Items or Requests
- ☐ d. Only Requests

Question 12

Complete

Mark 1.00 out of 1.00

Which setting enables a custom pipeline in Scrapy?

- ☐ a. MONGO_PIPELINE
- ☐ b. SPIDER_PIPELINE
- ☐ c. ITEM_STORAGE_BACKEND
- ☒ d. ITEM_PIPELINES

Question 13

Complete

Mark 1.00 out of 1.00

Why is dict(item) used before inserting into MongoDB?

- ☒ a. Items are not serializable directly
- ☐ b. To remove metadata
- ☐ c. MongoDB only accepts strings
- ☐ d. To convert the response to JSON

Question 14

Complete

Mark 1.00 out of 1.00

Why might this CSS selector return None? `quote.css('div.text::text').get()`

- ☐ a. The site uses JavaScript rendering
- ☒ b. Quotes use span.text, not div.text
- ☐ c. You forgot to install BeautifulSoup
- ☐ d. The element doesn't exist

Question 15

Complete

Mark 0.00 out of 1.00

You run `db.quotes.find({ rating: 'Five' }).count()` and get 0. What is a likely reason?

- ☐ a. Scrapy removes the rating before storing
- ☐ b. The rating field wasn't scraped properly or pipeline was skipped
- ☒ c. Ratings are stored as numbers
- ☐ d. You used the wrong MongoDB URI

