

Web Scraping Project Documentation

Overview:

The project consists of 4 core elements: Crawler, Scheduler, Apis & Tests.

Crawler: It is a program written in python which crawls a website (<https://books.toscrape.com>) using the popular scraping library beautifulsoup.

Scheduler: It is a python program that runs continuously and runs the crawler after a certain time to find any new or modified data in the website. If it finds any the database (DB) is updated accordingly, and json and csv reports are generated after the scheduler has finished crawling the entire website.

Apis: They are codes written in python using FastAPI to query and retrieve book data stored in the DB.

Tests: They are used to test the apis and the crawler.

- Incode documentation is also provided.

Primary Technologies and Libraries used:

Crawler: beautifulsoup

Scheduler: apscheduler

Apis: FastAPI

Tests: pytest

Database: MongoDB (MongoDB Compass)

Detailed Sections of The Project

Crawler:

This module contains an asynchronous web crawler designed to scrape books.toscrape.com (or any similar site) and store book metadata, snapshots, and change logs into a MongoDB database.

It handles:

- Fetching all paginated book links

- Scraping book details
- Detecting content changes via hashing
- Storing snapshots only when content changes
- Parallel crawling with concurrency limits
- Returning a list of all changes detected on a run

Environment Variables

- **BASE_URL** – Base website to crawl
- **CRAWL_CONCURRENCY** – Number of simultaneous requests
- **CRAWL_RETRIES** – Number of retry attempts for failed requests

Defaults:

- BASE_URL = "https://books.toscrape.com"
- CONCURRENCY = 10
- RETRIES = 3

Main Class: Crawler

`__init__(base_url, concurrency)`

Initializes the crawler with:

- Base URL
- Async HTTP client
- Semaphore to limit concurrency

`async fetch(url)`

Fetches HTML from a URL with:

- Concurrency limit
- Automatic retries
- Error logging

Returns: **HTML string**

`async get_all_book_links()`

Scrapes *every page* of the catalog to collect all book detail URLs.

Workflow:

1. Start at page 1
2. Extract <a> links for all books
3. Follow the **Next** button until pages end
4. Deduplicate links while keeping order

Returns: **List of book detail URLs**

`parse_rating(soup)`

Extracts star rating based on CSS class:
"One" → 1, "Two" → 2, ...

Returns: **Integer or None**

`parse_book(html, url)`

Parses a single book page and extracts:

- Title
- Description
- Category
- Prices (incl/excl tax)
- Availability
- Number of reviews
- Image URL
- Rating
- Unique book ID (derived from URL path)
- Timestamp
- Content hash

Returns:

`(book_data_dict, raw_html)`

`async process_book(url)`

Processes a single book.

Behavior:

If book exists in DB

- Compares stored hash vs. scraped hash
- If changed:
 - Create snapshot
 - Update book record
 - Insert change log
 - Return change info
- If unchanged:
 - No snapshot created
 - Returns None

If book is new

- Create snapshot
- Insert book
- Insert change log
- Return change info

Returns: **Change record or None**

`async run(resume=False)`

Main crawling workflow:

1. Get all book links
2. Process each book concurrently
3. Collect change records

Returns: **List of changes detected this run**

Helper Functions

`async crawl_all_books()`

Utility for schedulers:

1. Runs the crawler
2. Fetches all books from DB (without raw HTML)
3. Returns them as a list

Summary of Data Saved

Books Collection

- Metadata fields (title, category, price, etc.)
- content_hash
- raw_snapshot_id (link to snapshots collection)

Snapshots Collection

- Full HTML
- Timestamp
- Source URL

Change Log Collection

- book_id
- change_type (new / updated)
- recent (new / old)
- timestamps
- old vs new hash (for updates)

Book Model Using Pydantic

```
class Book(BaseModel):
    id: str = Field(..., description="Unique id")
    title: str
    description: Optional[str]
    category: Optional[str]
    price_including_tax: Optional[float]
    price_excluding_tax: Optional[float]
    availability: Optional[str]
    num_reviews: Optional[int]
    image_url: Optional[HttpUrl]
    rating: Optional[int] # 1-5
    source_url: HttpUrl
    crawl_timestamp: datetime
    content_hash: str
    raw_snapshot_id: Optional[str]
```

Schedular:

This module manages **automated crawling** of the book website using APScheduler. It triggers the crawler periodically, detects changes, and generates a daily report from those changes.

Logging Setup

A logger named "scheduler" is configured to print timestamped logs to the console to help track job execution and errors.

Function: `scheduled_crawl()`

An asynchronous task executed by the APScheduler.

Workflow:

1. Logs start of scheduled crawl
2. Creates a Crawler instance
3. Runs the crawler using `c.run()`
 - Returns a list of **new or updated books** detected during this run
4. Logs how many changes were detected
5. Calls `generate_daily_report(changes)`
 - Generates a report *only* from this run's changes
6. Closes the crawler's HTTP client

Returns:

Nothing, but generates a report and updates the DB.

Function: `async_main()`

Sets up and starts the APScheduler.

Behavior:

- Creates an AsyncIOScheduler for async tasks

Adds a job:

```
scheduler.add_job(scheduled_crawl, "interval", hours=24)
```

- → Runs `scheduled_crawl()` **every 24 hours**
- Starts the scheduler
- Keeps the program alive indefinitely using `await asyncio.Event().wait()`

Script Entry Point

If the file is executed directly:

```
asyncio.run(async_main())
```

This launches the scheduler, which will then execute the crawl job every 24 hours.

Summary

- `scheduled_crawl()`
→ Runs crawler → gets changes → generates report
- `async_main()`
→ Starts APScheduler → schedules the job → keeps program running
- APScheduler interval: **24 hours**

Alert & Report:

These modules are responsible for creating and sending reports respectively, after the scheduler crawls the entire website at a defined time interval.

Alert

The alert.py file is responsible for sending an email with the generated reports after each scheduler crawl. For testing purposes, an email is sent regardless of whether there are any changes or not.

Report

The report.py file generates csv & json reports every time the scheduler crawls the website. The files contain any change in data of the website or no change message.

APIs:

This module defines the **FastAPI backend** for the Book Scraper system.

It exposes endpoints to retrieve books, book details, and change logs.

It also enforces API key authentication, rate limiting, pagination, and optional filtering.

App Initialization

- Loads environment variables (e.g., API_PORT)

- Creates a FastAPI app with metadata
- Registers global rate limiting via SlowAPI
- Enables CORS for all origins (useful for frontend access)

API Key Authentication

All endpoints include:

```
dependencies=[Depends(get_api_key)]
```

Only clients providing a valid API key can access the API.

Rate Limiting

Each route uses:

```
@limiter.limit("X/hour")
```

Examples:

- /books: **100 requests/hour**
- /books/{book_id}: **100 requests/hour**
- /changes: **100 requests/hour**

This protects the API from abuse.

```
book_doc_to_resp(doc)
```

A helper function that extracts only the allowed fields from a MongoDB document before returning it to the client.

It prevents leaking internal fields like raw_snapshot_id or content_hash.

Endpoint: GET /books

Returns a **paginated list of books**, with optional filters and sorting.

Supported Query Parameters:

- category — filter by book category
- min_price, max_price — price range filter
- rating — filter by rating (1–5)

- `sort_by` — one of:
 - "rating" (desc)
 - "price" (asc)
 - "reviews" (desc)
- `page, page_size` — pagination controls

Example Response Body:

```
{
  "page": 1,
  "page_size": 20,
  "total": 1000,
  "results": [ ...list of books... ]
}
```

The query is translated into a MongoDB filter, sorted if needed, then paginated manually using `skip()` and `limit()`.

Endpoint: GET /books/{book_id}

Returns **one book** by its ID.

- Looks up `_id` in MongoDB
- If not found → returns 404 Book not found
- Otherwise → returns sanitized book document

Endpoint: GET /changes

Returns recent changes (new or updated books) from the `change_log` collection.

Details:

- Sorted by newest first
- Limited by query parameter `limit` (default 100)
- Converts `_id` and `book_id` to strings for JSON compatibility

Response:

```
{
  "results": [
    {
      "_id": "log123",
      "
```

```
        "book_id": "bk021",
        "change_type": "updated",
        "recent": "new",
        "timestamp": "...",
        "old_hash": "...",
        "new_hash": "..."
    }
]
}
```

Running the API

Use the following command:

```
uvicorn api.main:app --host 0.0.0.0 --port <API_PORT> --reload
```

Rate Limiter:

This module sets up **rate limiting** for the FastAPI application using **SlowAPI**.

limiter

```
limiter = Limiter(key_func=get_remote_address)
```

- Creates a global Limiter instance
- `key_func=get_remote_address` → limits requests per client IP address
- Can be used as a decorator on FastAPI routes to enforce rate limits, e.g.:

```
@limiter.limit("100/hour")
```

register_rate_limit(app: FastAPI)

Registers the Limiter with the FastAPI app.

Actions:

1. Attaches the limiter to `app.state.limiter`
2. Registers a **429 Too Many Requests** exception handler to return a proper response when a client exceeds the limit.

Usage:

```
from fastapi import FastAPI
from api.rate_limit import register_rate_limit

app = FastAPI()
register_rate_limit(app)
```

After this, any route decorated with `@limiter.limit()` is enforced, and exceeding clients get a 429 response.

Summary

- Provides a **global rate limiter** for the API
- Limits are **per client IP**
- Handles **429 errors gracefully**
- Works seamlessly with FastAPI decorators

Tests:

conftest.py

This module provides **fixtures and fake database classes** for testing the FastAPI Books Scraper API without requiring a real MongoDB instance. It enables **unit and integration tests** of API endpoints in isolation.

Key Components

1. Fake MongoDB Classes

- **FakeCursor**
Simulates a Motor cursor with support for `sort()`, `skip()`, `limit()`, and `to_list()`.
- **FakeCollection**
Simulates a MongoDB collection with basic methods:
 - `find_one()`, `find()`
 - `insert_one()`, `update_one()`
 - `count_documents()`
- Supports simple filtering, range queries (e.g., `$gte`, `$lte`), and sorting.
- **FakeDB**
Contains two collections:
 - `books` → stores book documents

- `change_log` → stores book change records

These classes allow testing database-dependent API logic without a real database.

2. Pytest Fixtures

- `sample_books` → Returns a list of sample book documents for testing.
- `fake_db` → Returns a FakeDB instance preloaded with sample books and change logs.
- `client` → Provides an AsyncClient for testing the FastAPI app with:
 - `get_db` patched to return `fake_db`
 - API key dependency overridden to allow "testapikey"
 - Uses ASGITransport to make requests to the app in-memory

Purpose

This setup allows writing **async API tests** that:

- Query books with filters, sorting, and pagination
- Retrieve individual book details
- Fetch change logs
- Test authentication and rate-limiting logic

All without touching a real database or network.

`test_api.py`

This module contains **asynchronous pytest tests** for the FastAPI Books Scraper API. It uses the client fixture (with a fake database and overridden API key) to test endpoints **without a real database or network**.

Tested Endpoints

1. **GET /books** – List books with optional filters, sorting, and pagination
2. **GET /books/{book_id}** – Retrieve a single book by ID
3. **GET /changes** – Retrieve recent changes (new or updated books)
4. **Authentication and rate limiting**

Tests Overview

Test	Purpose
<code>test_list_books_no_filters</code>	Returns all books when no filters are applied

test_list_books_category_filter	Returns only books in a specific category
test_list_books_price_range	Returns books within a specified price range
test_list_books_sorting_and_pagination	Validates sorting by rating and pagination
test_get_book_by_id_found	Returns correct book when ID exists
test_get_book_by_id_not_found	Returns 404 when book ID does not exist
test_changes_endpoint_default_limit	Returns all changes with default limit
test_changes_endpoint_with_limit	Returns only the specified number of change entries
test_unauthorized_access	Verifies 401 Unauthorized when API key is missing
test_validation_error_for_bad_query_param	Verifies 422 response for invalid query parameters
test_rate_limit_hit	Simulates rate limiting, expecting 429 if limit is exceeded

Key Features

- Uses **async tests** (@pytest.mark.asyncio) with AsyncClient
- Validates **response status codes**, payload structure, and content
- Covers **filtering, sorting, pagination, and authentication**
- Tests **rate limiting behavior** using SlowAPI configuration
- Fully isolated with **fake database and API key override**

Summary

These tests ensure that:

1. API endpoints behave correctly with various query parameters.
2. Authentication and rate limiting are enforced.
3. Edge cases (missing book, invalid input, query limits) are handled gracefully.

The test suite can be run with:

pytest tests/test_api.py

or,

pytest -v

test_crawler.py

This module contains **asynchronous tests** for the Crawler class in the Books Scraper system.

Tested Functionality

get_all_book_links()

- Ensures the crawler correctly **collects all book detail URLs** from paginated catalogue pages.
- Uses **monkeypatching** to simulate HTTP responses, avoiding actual network requests.

Test Overview

Test	Purpose
test_get_all_book_links_monkeypatched	Simulates two catalogue pages with next links and multiple books. Verifies that Crawler.get_all_book_links(): <ul style="list-style-type: none">• Follows pagination• Extracts all book URLs• Preserves correct order

Key Features

- **monkeypatch** replaces Crawler.fetch with fake_fetch returning predefined HTML.
- Tests **HTML parsing, pagination, and URL joining** without external dependencies.
- Verifies that the crawler correctly returns all expected links.

Summary

This test ensures the crawler:

1. Handles multi-page catalogues.
2. Extracts book URLs from the HTML correctly.
3. Returns a deduplicated, ordered list of links.

Swagger UI for API testing:

- To use the UI for testing go to the following url (after turning on the server):
<http://127.0.0.1:8000/docs>

- After going to the UI click on the authorize button to enter the authorization header.

The screenshot shows the Books Scraper API documentation at 127.0.0.1:8000/docs. At the top right, there is a green 'Authorize' button with a lock icon. A red arrow points from the text above to this button. Below the button, the 'default' section lists three API endpoints: GET /books (List Books), GET /books/{book_id} (Get Book), and GET /changes (Get Changes). Each endpoint has a lock icon to its right. At the bottom, there are two sections for 'Schemas': 'HTTPValidationError' and 'ValidationErrorResponse', each with a 'Expand all object' link.

- Then input the value “testapikey” as the authentication header value like in the picture below:

The screenshot shows the same API documentation page with an open 'Available authorizations' dialog. This dialog is titled 'Available authorizations' and contains one entry: 'APIKeyHeader (apiKey)'. It shows the 'Name: X-API-Key', 'In: header', and 'Value: testapikey'. Below the dialog are the familiar 'Authorize' and 'Close' buttons. The background shows the same API endpoints and schema sections as the first screenshot.

- Now you can test the APIs.

Important Screenshots:

- Crawler starting to crawl the website.

```

File Edit Selection View Go Run Terminal Help ← → ⌘ Web Crawler
EXPLORER PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
WEB CRAWLER
> venv
> api
> __pycache__
> __init__.py
auth.py
main.py
rate_limit.py
> crawler
> __pycache__
> __init__.py
crawler.py
db.py
models.py
utils.py
> reports
scheduler
> __pycache__
__init__.py
reporter.py
scheduler.py
tests
> __pycache__
confest.py
test_api.py
test_crawler.py
utils
.env
pytest.ini
requirements.txt
OUTLINE
TIMELINE
Ln 185, Col 74 Spaces: 4 UTF-8 CRLF {} Python ⌘ 3.13.3 (venv) ⌘ Go Live ⌘ Prettier ⌘
80°F Sunny

```

2) Crawler inserting books into MongoDB.

```

File Edit Selection View Go Run Terminal Help ← → ⌘ Web Crawler
EXPLORER PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
WEB CRAWLER
2025-12-01 13:30:14,426 INFO Inserted new book of-mice-and-men_37
2025-12-01 13:30:14,427 INFO Inserted new book off-sides-off_1_38
2025-12-01 13:30:14,453 INFO Inserted new book myriad-prentor_1_36
2025-12-01 13:30:14,454 INFO Inserted new book paradise-lost-paradise_1_45
2025-12-01 13:30:14,512 INFO Inserted new book ms-marvel-vol-1-no-normal-ms-marvel-2014-2015_1_34
2025-12-01 13:30:14,530 INFO Inserted new book matilda_32
2025-12-01 13:30:14,589 INFO Inserted new book lost-among-the-living_31
2025-12-01 13:30:14,601 INFO Inserted new book old-records-never-die-one-mans-quest-for-his-vinyl-and-his-past_39
2025-12-01 13:30:14,747 INFO Inserted new book lord-of-the-flies_30
2025-12-01 13:30:14,747 INFO Inserted new book listen-to-me-fusion_1_29
2025-12-01 13:30:14,748 INFO Inserted new book kitchens-of-the-great-midwest_28
2025-12-01 13:30:14,803 INFO Inserted new book imperfect-harmony_26
2025-12-01 13:30:14,825 INFO Inserted new book icing-aces-hockey_2_25
2025-12-01 13:30:14,825 INFO Inserted new book hawkeye-vol-1-my-life-as-a-weapon-hawkeye_1_24
2025-12-01 13:30:14,885 INFO Inserted new book having-the-barbarians-baby-ice-planet-barbarians_75_23
2025-12-01 13:30:14,897 INFO Inserted new book giant-days-vol-1-giant-days_1-4_22
2025-12-01 13:30:15,057 INFO Inserted new book fruits-basket-vol-1-fruits-basket_1_21
2025-12-01 13:30:15,058 INFO Inserted new book jane-eye_27
2025-12-01 13:30:15,059 INFO Inserted new book frankenstein_20
2025-12-01 13:30:15,059 INFO Inserted new book forever-rockers-the-rocker_12_19
2025-12-01 13:30:15,119 INFO Inserted new book fighting-fate-fighting_6_18
2025-12-01 13:30:15,120 INFO Inserted new book emma_17
2025-12-01 13:30:15,120 INFO Inserted new book eat-pray-love_16
2025-12-01 13:30:15,176 INFO Inserted new book deep-under-walker-security_1_15
2025-12-01 13:30:15,333 INFO Inserted new book charlie-and-the-chocolate-factory-charlie-bucket_1_13
2025-12-01 13:30:15,360 INFO Inserted new book bridget-jones-diary-bridget-jones_1_10
2025-12-01 13:30:15,360 INFO Inserted new book bright-lines_11
2025-12-01 13:30:15,393 INFO Inserted new book bounty-colorado-mountain_7_9
2025-12-01 13:30:15,431 INFO Inserted new book blood-defense-samantha-brinkman_1_8
2025-12-01 13:30:15,432 INFO Inserted new book bleach-vol-1-strawberry-and-the-soul-reapers-bleach_1_7
2025-12-01 13:30:15,433 INFO Inserted new book my-perfect-mistake-over-the-top_1_35
2025-12-01 13:30:15,477 INFO Inserted new book beyond-good-and-evil_6
2025-12-01 13:30:15,477 INFO Inserted new book choosing-our-religion-the-spiritual-lives-of-americas-nones_14
2025-12-01 13:30:15,627 INFO Inserted new book alice-in-wonderland-alices-adventures-in-wonderland_1_12
2025-12-01 13:30:15,638 INFO Inserted new book charity-cross-charles-towne-bellies_4_12
2025-12-01 13:30:15,651 INFO Inserted new book ojin-dem-human-volume-1-ajin-dem-human_1_4
2025-12-01 13:30:15,654 INFO Inserted new book a-spys-devotion-the-regency-spies-of-london_1_3
2025-12-01 13:30:15,687 INFO Inserted new book 1st-to-die-womens-murder-club_1_2
2025-12-01 13:30:15,697 INFO Inserted new book 1000-places-to-see-before-you-die_1
Ln 185, Col 74 Spaces: 4 UTF-8 CRLF {} Python ⌘ 3.13.3 (venv) ⌘ Go Live ⌘ Prettier ⌘
80°F Sunny

```

3) Crawler crawled the website and started to insert books into MongoDB.

```

File Edit Selection View Go Run Terminal Help ← → ⌘ Web Crawler
EXPLORER PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
WEB CRAWLER
> venv
2025-12-01 13:29:28,312 INFO Listing page: https://books.toscrape.com/catalogue/page-22.html
2025-12-01 13:29:28,626 INFO Listing page: https://books.toscrape.com/catalogue/page-23.html
2025-12-01 13:29:29,269 INFO Listing page: https://books.toscrape.com/catalogue/page-24.html
2025-12-01 13:29:29,589 INFO Listing page: https://books.toscrape.com/catalogue/page-25.html
2025-12-01 13:29:29,916 INFO Listing page: https://books.toscrape.com/catalogue/page-26.html
2025-12-01 13:29:30,289 INFO Listing page: https://books.toscrape.com/catalogue/page-27.html
2025-12-01 13:29:30,891 INFO Listing page: https://books.toscrape.com/catalogue/page-28.html
2025-12-01 13:29:31,238 INFO Listing page: https://books.toscrape.com/catalogue/page-29.html
2025-12-01 13:29:31,564 INFO Listing page: https://books.toscrape.com/catalogue/page-30.html
2025-12-01 13:29:31,893 INFO Listing page: https://books.toscrape.com/catalogue/page-31.html
2025-12-01 13:29:32,219 INFO Listing page: https://books.toscrape.com/catalogue/page-32.html
2025-12-01 13:29:32,525 INFO Listing page: https://books.toscrape.com/catalogue/page-33.html
2025-12-01 13:29:32,842 INFO Listing page: https://books.toscrape.com/catalogue/page-34.html
2025-12-01 13:29:33,162 INFO Listing page: https://books.toscrape.com/catalogue/page-35.html
2025-12-01 13:29:33,494 INFO Listing page: https://books.toscrape.com/catalogue/page-36.html
2025-12-01 13:29:33,833 INFO Listing page: https://books.toscrape.com/catalogue/page-37.html
2025-12-01 13:29:34,149 INFO Listing page: https://books.toscrape.com/catalogue/page-38.html
2025-12-01 13:29:34,458 INFO Listing page: https://books.toscrape.com/catalogue/page-39.html
2025-12-01 13:29:34,768 INFO Listing page: https://books.toscrape.com/catalogue/page-40.html
2025-12-01 13:29:35,076 INFO Listing page: https://books.toscrape.com/catalogue/page-41.html
2025-12-01 13:29:35,384 INFO Listing page: https://books.toscrape.com/catalogue/page-42.html
2025-12-01 13:29:35,692 INFO Listing page: https://books.toscrape.com/catalogue/page-43.html
2025-12-01 13:29:36,013 INFO Listing page: https://books.toscrape.com/catalogue/page-44.html
2025-12-01 13:29:36,343 INFO Listing page: https://books.toscrape.com/catalogue/page-45.html
2025-12-01 13:29:36,663 INFO Listing page: https://books.toscrape.com/catalogue/page-46.html
2025-12-01 13:29:37,289 INFO Listing page: https://books.toscrape.com/catalogue/page-47.html
2025-12-01 13:29:37,891 INFO Listing page: https://books.toscrape.com/catalogue/page-48.html
2025-12-01 13:29:38,486 INFO Listing page: https://books.toscrape.com/catalogue/page-49.html
2025-12-01 13:29:39,078 INFO Listing page: https://books.toscrape.com/catalogue/page-50.html
2025-12-01 13:29:39,688 INFO Found 1000 book links
2025-12-01 13:29:39,976 INFO Inserted new book a-light-in-the-attic_1000
2025-12-01 13:29:40,282 INFO Inserted new book starving-hearts-triangular-trade-trilogy_1990
2025-12-01 13:29:40,583 INFO Inserted new book shakespeares-sonnets_989
2025-12-01 13:29:40,845 INFO Inserted new book the-requiem-red_995
2025-12-01 13:29:40,846 INFO Inserted new book the-black-maria_991
2025-12-01 13:29:40,910 INFO Inserted new book the-boys-in-the-boat-nine-americans-and-their-epic-quest-for-gold-at-the-1936-berlin-olympics_992
2025-12-01 13:29:40,912 INFO Inserted new book the-coming-woman-a-novel-based-on-the-life-of-the-infamous-feminist-victoria-woodhull_993
2025-12-01 13:29:40,913 INFO Inserted new book sapiens-a-brief-history-of-humankind_996
2025-12-01 13:29:40,913 INFO Inserted new book sumission_998
2025-12-01 13:29:40,916 INFO Inserted new book tipping-the-velvet_999
2025-12-01 13:29:40,916 INFO Inserted new book set-me-free_988

```

Ln 185, Col 74 Spaces: 4 UTF-8 CRLF {} Python 3.13.3 (venv) Go Live Prettier 1:31 PM 12/1/2025

4) Scheduler starting to crawl.

```

File Edit Selection View Go Run Terminal Help ← → ⌘ Web Crawler
EXPLORER PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
WEB CRAWLER
> venv
(venv) PS D:\Codes\Python\Web Crawler> python -m scheduler.scheduler
2025-12-01 16:13:12,710 [INFO] Scheduler started (daily at 02:00 UTC)
2025-12-01 16:18:12,746 [INFO] Starting scheduled crawl
2025-12-01 16:18:13,472 INFO Listing page: https://books.toscrape.com/catalogue/page-1.html
2025-12-01 16:18:21,764 INFO Listing page: https://books.toscrape.com/catalogue/page-2.html
2025-12-01 16:18:25,297 INFO Listing page: https://books.toscrape.com/catalogue/page-3.html
2025-12-01 16:18:26,669 INFO Listing page: https://books.toscrape.com/catalogue/page-4.html
2025-12-01 16:18:28,137 INFO Listing page: https://books.toscrape.com/catalogue/page-5.html
2025-12-01 16:18:29,553 INFO Listing page: https://books.toscrape.com/catalogue/page-6.html
2025-12-01 16:18:32,330 INFO Listing page: https://books.toscrape.com/catalogue/page-7.html
2025-12-01 16:18:35,437 INFO Listing page: https://books.toscrape.com/catalogue/page-8.html
2025-12-01 16:18:38,161 INFO Listing page: https://books.toscrape.com/catalogue/page-9.html
2025-12-01 16:18:41,902 INFO Listing page: https://books.toscrape.com/catalogue/page-10.html
2025-12-01 16:18:46,017 INFO Listing page: https://books.toscrape.com/catalogue/page-11.html
2025-12-01 16:18:50,318 INFO Listing page: https://books.toscrape.com/catalogue/page-12.html

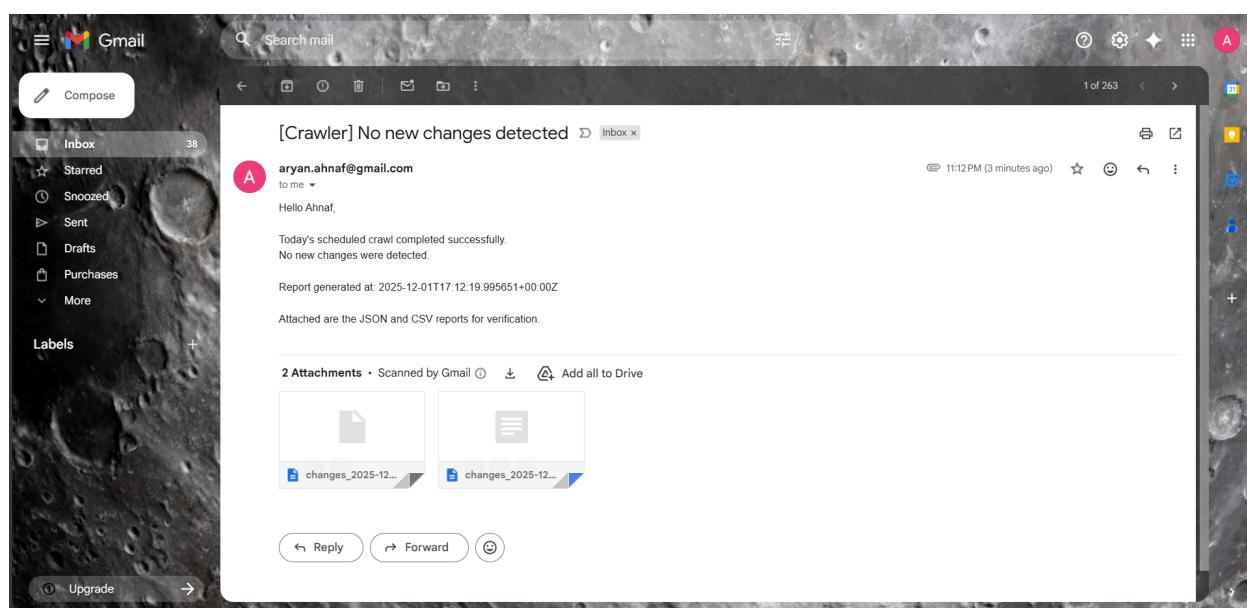
```

Sunset coming 5:10 PM

Ln 163, Col 41 Spaces: 4 UTF-8 CRLF {} Python 3.13.3 (venv) Go Live Prettier 4:18 PM 12/1/2025

5) The scheduler finished crawling and comparing the hashes of each book to detect any changes and then generated reports and sent an email alert.

The screenshot shows a Microsoft Visual Studio Code (VS Code) interface. The top bar includes standard menu items like File, Edit, Selection, View, Go, Run, Terminal, Help, and a search bar labeled 'Web Crawler'. The left sidebar has sections for 'EXPLORER', 'PROBLEMS', 'OUTPUT', 'DEBUG CONSOLE', 'TERMINAL', and 'PORTS'. The 'EXPLORER' section displays a tree view of files and folders, including 'WEB CRAWLER', 'api', 'crawler', 'reports', 'changes.2025-12-01.csv', 'changes.2025-12-01.json', 'scheduler', 'tests', 'utils', and 'OUTLINE'. The main workspace shows a large list of log entries from a crawler, starting with '2025-12-01 23:12:17,445 INFO No changes for book of mice-and-men_37 (content hash matched)' and ending with '2025-12-01 23:12:19,988 [INFO] Scheduled crawl finished, 0 changes detected'. The bottom right corner shows the status bar with '112.12 PM', battery level '70%', and connectivity icons.



6) All 12 tests passed.

```
(.venv) PS D:\Codes\Python\Web Crawler> pytest -v
=====
platform win32 -- Python 3.13.3, pytest-9.0.1, pluggy-1.6.0
rootdir: D:\Codes\Python\Web Crawler
configfile: pytest.ini
plugins: anyio-4.11.0, asyncio-1.3.0
asyncio: mode=Mode.AUTO, debug=False, asyncio_default_fixture_loop_scope=None, asyncio_default_test_loop_scope=function
collected 12 items

tests\test_api.py ..... 12 passed in 0.54s
```

API testing using an API Client:

- 1) <http://localhost:8000/books>

THUNDER CLIENT

New Request

Activity Collections Env

filter activity

GET localhost:8000/books 5 hours ago

HTTP Headers

- Accept */*
- User-Agent Thunder Client (https://www.thunderclient.com)
- X-API-Key testapikey
- header value

Status: 200 OK Size: 39.39 KB Time: 88 ms

```

1  {
2    "page": 1,
3    "page_size": 20,
4    "total": 1000,
5    "results": [
6      {
7        "_id": "a-light-in-the-attic_1000",
8        "title": "A Light in the Attic",
9        "description": "It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love them. It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love them. Need proof of his genius? RockabyeRockabye baby, in the treetopon't you know a treetop no safe place to rock? And who put you up there, And your cradle, too? Baby, I think someone down here's got it in for you. Shel, you never sounded so good...more",
10       "category": "Poetry",
11       "price_including_tax": 51.77,
12       "price_excluding_tax": 51.77,
13       "availability": "In stock (22 available)",
14       "num_reviews": 0,

```

Response Chart ↻

Go Live

11:11 PM 11/29/2025

2) http://localhost:8000/books?category=Mystery

THUNDER CLIENT

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

HTTP Headers

- Accept */*
- User-Agent Thunder Client (https://www.thunderclient.com)
- X-API-Key testapikey
- header value

Status: 200 OK Size: 34.69 KB Time: 49 ms

```

1  {
2    "page": 1,
3    "page_size": 20,
4    "total": 32,
5    "results": [
6      {
7        "_id": "sharp-objects_997",
8        "title": "Sharp Objects",
9        "description": "WICKED above her hipbone, GIRL across her heart Words are like a road map to reporter Camille Preacher's troubled past. Fresh from a brief stay at a psych hospital, Camille's first assignment from the second-rate daily paper where she works brings her reluctantly back to her hometown to cover the murders of two preteen girls. NASTY on her kneecap, BABYDOLL on her leg Since WICKED above her hipbone, GIRL across her heart Words are like a road map to reporter Camille Preacher's troubled past. Fresh from a brief stay at a psych hospital, Camille's first assignment from the second-rate daily paper where she works brings her reluctantly back to her hometown to cover the murders of two preteen girls. NASTY on her kneecap, BABYDOLL on her leg Since she left town eight years ago, Camille has hardly spoken to her neurotic, hypochondriac mother or to the half-sister she barely knows: a beautiful thirteen-year-old with an eerie grip on the town. Now, installed again in her family's Victorian mansion, Camille is haunted by the childhood tragedy she has spent her whole life trying to cut from her memory. HARMFUL on her wrist, WHORE on her ankle As Camille works to uncover the truth about these violent crimes, she finds herself identifying with the young victims—a bit too strongly. Clues keep leading to dead ends, forcing Camille to unravel the

```

Copy Response Chart ↻

Go Live

11:11 PM 11/29/2025

3) http://localhost:8000/books?rating=4

The screenshot shows the Thunder Client interface with the following details:

- Request URL:** GET http://127.0.0.1:8000/books?rating=4
- Request Headers:**
 - Accept: */*
 - User-Agent: Thunder Client (https://www.thunderclient.com)
 - X-API-Key: testapikey
 - header: value
- Response Status:** 200 OK
- Response Size:** 37.07 KB
- Response Time:** 41 ms
- Response Content:** A JSON object representing a book entry, including fields like id, title, description, category, price, availability, reviews, image URL, and rating.

4) http://localhost:8000/books?min_price=10

Thunder Client

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

HTTP Headers

Raw

Accept */*

User-Agent Thunder Client (https://www.thunderclient.com)

X-API-Key testapikey

header value

Status: 200 OK Size: 39.39 KB Time: 42 ms

Response Headers Cookies Results Docs

Copy

```
1 {  
2   "page": 1,  
3   "page_size": 20,  
4   "total": 1000,  
5   "results": [  
6     {  
7       "_id": "a-light-in-the-attic_1000",  
8       "title": "A Light in the Attic",  
9       "description": "It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love them. It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love that Silverstein. Need proof of his genius? RockabyeRockabye baby, in the treetopdon't you know a treetopis no safe place to rock?And who put you up there,And your cradle, too?Baby, I think someone down here'sgot it in for you. Shel, you never sounded so good....more",  
10      "category": "Poetry",  
11      "price_including_tax": 51.77,  
12      "price_excluding_tax": 51.77,  
13      "availability": "In stock (22 available)",  
14      "num_reviews": 0,  
15    }  
16  ]  
17}  
18
```

Response Chart

67°F Clear 11:12 11/29/2022

5) http://localhost:8000/books?max_price=20

```

Status: 200 OK  Size: 40.03 KB  Time: 10 ms

Response Headers:
Content-Type: application/json
Content-Length: 196
Date: Mon, 29 Nov 2025 22:13:25 GMT
Server: WSGI/1.0
Connection: keep-alive
Vary: Accept

HTTP Headers:
Accept: */*
User-Agent: Thunder Client (https://www.thunderclient.com)
X-API-Key: testapikey
header: value

Response Body:
{
    "page": 1,
    "page_size": 20,
    "total": 196,
    "results": [
        {
            "_id": "starving-hearts-triangular-trade-trilogy-1_990",
            "title": "Starving Hearts (Triangular Trade Trilogy, #1)",
            "description": "since her assault, Miss Annette Chetwynd has been plagued by nightmares and worries about an arranged marriage. But she yearns to find her anonymous rescuer. Unfortunately, her health and intellect prevent it. Both repel suitors and cause Annette to doubt God's existence, at least until He answers her prayers in an unusual way ... Mr. Peter Adsley is joining the clergy, an since her assault, Miss Annette Chetwynd has been plagued by nightmares and worries about an arranged marriage. But she yearns to find her anonymous rescuer. Unfortunately, her health and intellect prevent it. Both repel suitors and cause Annette to doubt God's existence, at least until He answers her prayers in an unusual way ... Mr. Peter Adsley is joining the clergy, and he desires a godly wife by his side. After his failed attempt to obtain one, he engages in a clandestine meeting with the bewitching young woman who keeps crossing his path. But she is so unsettling. Destined for disappointments, Peter and Annette endure disgraceful situations. Will Peter's faith sustain him through overwhelming setbacks? Can Annette overcome her doubts? Or will their starving hearts yield to the machinations of a mad man? ...more",
            "category": "Default",
            "price_including_tax": 13.99
        }
    ]
}

```

6) http://localhost:8000/books?min_price=10&max_price=20

```

Status: 200 OK  Size: 40.03 KB  Time: 11 ms

Response Headers:
Content-Type: application/json
Content-Length: 196
Date: Mon, 29 Nov 2025 22:13:25 GMT
Server: WSGI/1.0
Connection: keep-alive
Vary: Accept

HTTP Headers:
Accept: */*
User-Agent: Thunder Client (https://www.thunderclient.com)
X-API-Key: testapikey
header: value

Response Body:
{
    "page": 1,
    "page_size": 20,
    "total": 196,
    "results": [
        {
            "_id": "starving-hearts-triangular-trade-trilogy-1_990",
            "title": "Starving Hearts (Triangular Trade Trilogy, #1)",
            "description": "since her assault, Miss Annette Chetwynd has been plagued by nightmares and worries about an arranged marriage. But she yearns to find her anonymous rescuer. Unfortunately, her health and intellect prevent it. Both repel suitors and cause Annette to doubt God's existence, at least until He answers her prayers in an unusual way ... Mr. Peter Adsley is joining the clergy, an since her assault, Miss Annette Chetwynd has been plagued by nightmares and worries about an arranged marriage. But she yearns to find her anonymous rescuer. Unfortunately, her health and intellect prevent it. Both repel suitors and cause Annette to doubt God's existence, at least until He answers her prayers in an unusual way ... Mr. Peter Adsley is joining the clergy, and he desires a godly wife by his side. After his failed attempt to obtain one, he engages in a clandestine meeting with the bewitching young woman who keeps crossing his path. But she is so unsettling. Destined for disappointments, Peter and Annette endure disgraceful situations. Will Peter's faith sustain him through overwhelming setbacks? Can Annette overcome her doubts? Or will their starving hearts yield to the machinations of a mad man? ...more",
            "category": "Default",
            "price_including_tax": 13.99
        }
    ]
}

```

7) http://127.0.0.1:8000/books?category=Romance&min_price=15&max_price=30&rating=5

THUNDER CLIENT

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

HTTP Headers

Accept */*

User-Agent Thunder Client (https://www.thunderclient.com)

X-API-Key testapikey

header value

Status: 200 OK Size: 1.66 KB Time: 42 ms

Response Headers Cookies Results Docs

```
1 {  
2   "page": 1,  
3   "page_size": 20,  
4   "total": 1,  
5   "results": [  
6     {  
7       "id": "chase-me-paris-nights-2_977",  
8       "title": "Chase Me (Paris Nights #2)",  
9       "description": "A Michelin two-star chef at twenty-eight, Violette Lenoir could handle anything, including a cocky burglar who broke into her restaurant in the middle of the night. Or so she thought. Elite counterterrorist operative Chase \"Smith\" had been through things that made Hell Week look easy. But nothing had prepared him for a leather-clad blonde who held him at bay at knifepoint. An A Michelin two-star chef at twenty-eight, Violette Lenoir could handle anything, including a cocky burglar who broke into her restaurant in the middle of the night. Or so she thought. Elite counterterrorist operative Chase \"Smith\" had been through things that made Hell Week look easy. But nothing had prepared him for a leather-clad blonde who held him at bay at knifepoint and dared him to take her on. Now if only saving the world didn't require her to ruin her life. Two people who thought they could handle anything now have to take on each other. It's a battle neither one expected. But with their futures on the line, they have nothing to lose...but their hearts. Warning: This book contains one arrogant Navy SEAL, nights of Paris passion, and a woman who wants to have it all. ...more",  
10      "category": "Romance",  
11      "price_including_tax": 25.27,  
12      "price_excluding_tax": 25.27,  
13    }  
14  ]  
15 }  
16 }
```

Response Chart Go Live

11:13 PM 11/29/2025

8) http://127.0.0.1:8000/books?sort_by=price

THUNDER CLIENT

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

HTTP Headers

Accept */*

User-Agent Thunder Client (https://www.thunderclient.com)

X-API-Key testapikey

header value

Status: 200 OK Size: 34.14 KB Time: 11 ms

Response Headers Cookies Results Docs

```
1 {  
2   "page": 1,  
3   "page_size": 20,  
4   "total": 1000,  
5   "results": [  
6     {  
7       "id": "an-abundance-of-katherines_362",  
8       "title": "An Abundance of Katherines",  
9       "description": "Katherine V thought boys were gross. Katherine X just wanted to be friends. Katherine XVI ITT dumped him in an e-mail. K-19 broke his heart when it comes to relationships. Colin Singleton's type happens to be girls named Katherine. And when it comes to girls named Katherine, Colin is always getting dumped. Nineteen times, to be exact. On a road trip miles from home, this anagram-happy, washed-up child prodigy has ten thousand dollars in his pocket, a bloodthirsty feral hog on his trail, and an overweight, Judge Judy-loving best friend riding shotgun—but no Katherines. Colin is on a mission to prove The Theorem of Underlying Katherine Predictability, which he hopes will predict the future of any relationship, avenge dumpies everywhere, and finally win the girl. Love, friendship, and a dead Austro-Hungarian archduke add up to surprising and heart-changing conclusions in this ingeniously layered comic novel about reinventing oneself. ...more",  
10     }  
11   ]  
12 }  
13 }
```

Response Chart Go Live

11:15 PM 11/29/2025

9) http://127.0.0.1:8000/books?sort_by=rating

The screenshot shows the Thunder Client interface. A GET request is made to `http://127.0.0.1:8000/books?sort_by=rating`. The response is successful (200 OK), with a size of 36.4 KB and a duration of 10 ms. The response body is a JSON object representing a book:

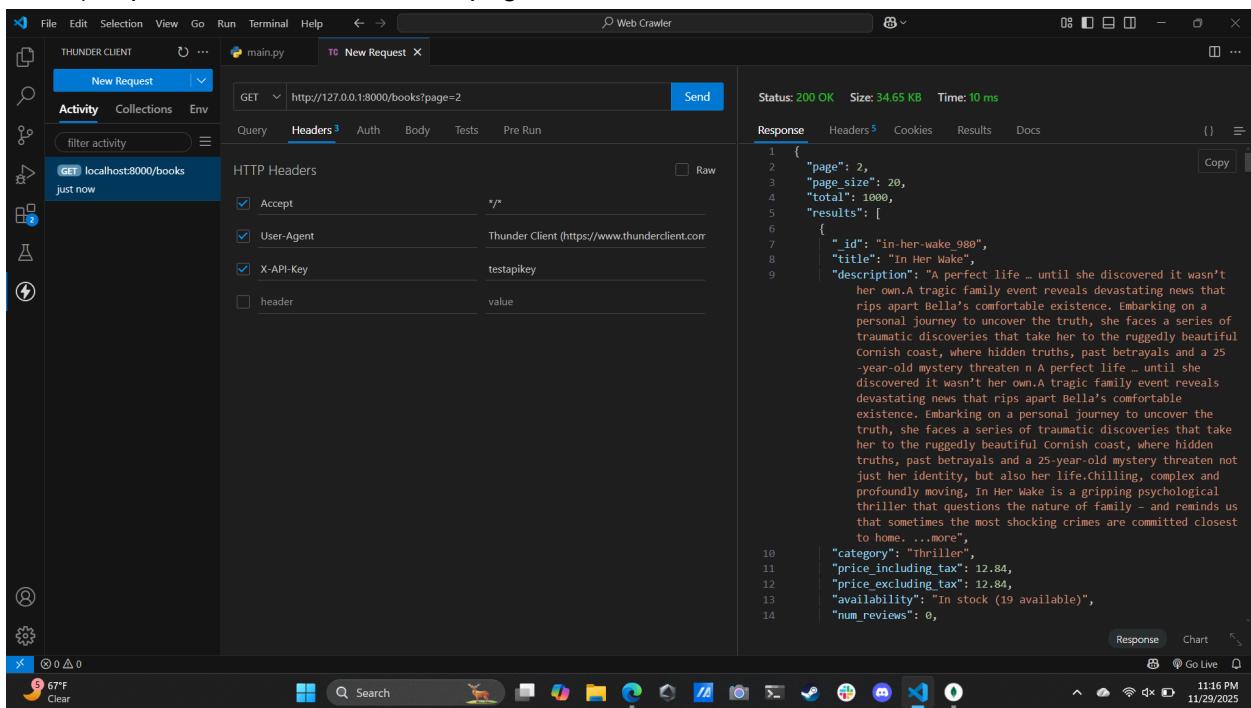
```
1 {
2     "page": 1,
3     "page_size": 20,
4     "total": 1000,
5     "results": [
6         {
7             "_id": "sapiens-a-brief-history-of-humankind_996",
8             "title": "Sapiens: A Brief History of Humankind",
9             "description": "From a renowned historian comes a groundbreaking narrative of humanity's creation and evolution—a #1 international bestseller—that explores the ways in which biology and history have defined us and enhanced our understanding of what it means to be \"human.\" One hundred thousand years ago, at least six different species of humans inhabited Earth. Yet today there is only one—us. From a renowned historian comes a groundbreaking narrative of humanity's creation and evolution—a #1 international bestseller—that explores the ways in which biology and history have defined us and enhanced our understanding of what it means to be \"human.\" One hundred thousand years ago, at least six different species of humans inhabited Earth. Yet today there is only one—us. Homo sapiens. What happened to the others? And what may happen to us? Most books about the history of humanity pursue either a historical or a biological approach, but Dr. Yuval Noah Harari breaks the mold with this highly original book that begins about 70,000 years ago with the appearance of modern cognition. From examining the role evolving humans have played in the global ecosystem to charting the rise of empires, Sapiens integrates history and science to reconsider accepted narratives, connect past developments with contemporary concerns, and"
10     ]
```

10) http://127.0.0.1:8000/books?sort_by=reviews

The screenshot shows the Thunder Client interface. A GET request is made to `http://127.0.0.1:8000/books?sort_by=reviews`. The response is successful (200 OK), with a size of 39.39 KB and a duration of 10 ms. The response body is a JSON object representing a book:

```
1 {
2     "page": 1,
3     "page_size": 20,
4     "total": 1000,
5     "results": [
6         {
7             "_id": "mesaerion-the-best-science-fiction-stories-1800-1849_983",
8             "title": "Mesaerion: The Best Science Fiction Stories 1800-1849",
9             "description": "Andrew Barger, award-winning author and engineer, has extensively researched forgotten journals and magazines of the early 19th century to locate groundbreaking science fiction short stories in the English language. In doing so, he found what is possibly the first science fiction story by a female (and it is not from Mary Shelley). Andrew located the first steampunk short story by a female (and it is not from Mary Shelley). Andrew located the first steampunk short story, which has not been republished since 1844. There is the first voyage to the moon in a balloon, republished for the first time since 1820 that further tells of a darkness machine and a lunarian named Zuloc. Other sci-stories include the first robotic insect and an electricity gun. Once again, Andrew has searched old texts to find the very best science fiction stories from the period when the genre automated to life, some of the stories are published for the first time in nearly 200 years. Read these fantastic stories today! OUR OWN"
10     ]
```

11) http://127.0.0.1:8000/books?page=2



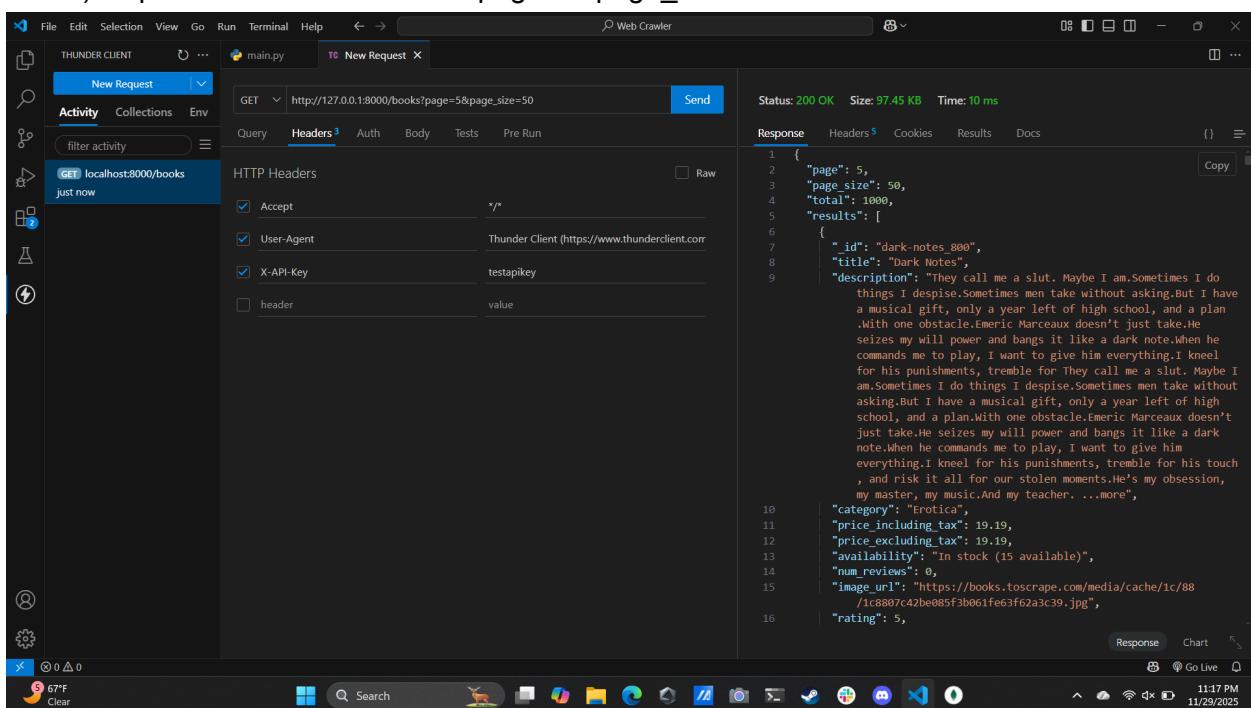
The screenshot shows the Thunder Client interface. On the left, there's a sidebar with icons for file operations, activity logs, collections, environment variables, and a search bar. The main area has tabs for 'New Request' and 'main.py'. A 'New Request' card is selected, showing a GET request to 'http://127.0.0.1:8000/books?page=2'. The 'Headers' tab is active, displaying the following headers:

```
Accept: */*
User-Agent: Thunder Client (https://www.thunderclient.com)
X-API-Key: testapikey
```

The 'Response' tab shows the API response:

```
1 {
2     "page": 2,
3     "page_size": 20,
4     "total": 1000,
5     "results": [
6         {
7             "_id": "in-her-wake_980",
8             "title": "In Her Wake",
9             "description": "A perfect life ... until she discovered it wasn't her own. A tragic family event reveals devastating news that rips apart Bella's comfortable existence. Embarking on a personal journey to uncover the truth, she faces a series of traumatic discoveries that take her to the ruggedly beautiful Cornish coast, where hidden truths, past betrayals and a 25-year-old mystery threaten n A perfect life ... until she discovered it wasn't her own. A tragic family event reveals devastating news that rips apart Bella's comfortable existence. Embarking on a personal journey to uncover the truth, she faces a series of traumatic discoveries that take her to the ruggedly beautiful Cornish coast, where hidden truths, past betrayals and a 25-year-old mystery threaten not just her identity, but also her life. Chilling, complex and profoundly moving, In Her Wake is a gripping psychological thriller that questions the nature of family - and reminds us that sometimes the most shocking crimes are committed closest to home. ...more",
10            "category": "thriller",
11            "price_including_tax": 12.84,
12            "price_excluding_tax": 12.84,
13            "availability": "In stock (19 available)",
14            "num_reviews": 0,
```

12) http://127.0.0.1:8000/books?page=5&page_size=50



The screenshot shows the Thunder Client interface. The layout is identical to the previous screenshot, with the sidebar, 'New Request' card, and 'main.py' tab. The 'New Request' card now shows a GET request to 'http://127.0.0.1:8000/books?page=5&page_size=50'. The 'Headers' tab is active, displaying the same set of headers as the previous request.

The 'Response' tab shows the API response:

```
1 {
2     "page": 5,
3     "page_size": 50,
4     "total": 1000,
5     "results": [
6         {
7             "_id": "dark-notes_800",
8             "title": "Dark Notes",
9             "description": "They call me a slut. Maybe I am. Sometimes I do things I despise. sometimes men take without asking. But I have a musical gift, only a year left of high school, and a plan .With one obstacle.Emeric Marceaux doesn't just take. He seizes my will power and bangs it like a dark note. When he commands me to play, I want to give him everything. I kneel for his punishments, tremble for They call me a slut. Maybe I am. Sometimes I do things I despise. Sometimes men take without asking. But I have a musical gift, only a year left of high school, and a plan. With one obstacle.Emeric Marceaux doesn't just take. He seizes my will power and bangs it like a dark note. When he commands me to play, I want to give him everything. I kneel for his punishments, tremble for his touch , and risk it all for our stolen moments. He's my obsession, my master, my music. And my teacher. ...more",
10            "category": "Erotica",
11            "price_including_tax": 19.19,
12            "price_excluding_tax": 19.19,
13            "availability": "In stock (15 available)",
14            "num_reviews": 0,
15            "image_url": "https://books.toscrape.com/media/cache/1c/88/1c8807c42be085f3b061fe63f62a3c39.jpg",
16            "rating": 5,
```

13) http://127.0.0.1:8000/books?category=Fantasy&min_price=5&max_price=40&rating=4&sort_by=rating&page=2&page_size=30

THUNDER CLIENT

File Edit Selection View Go Run Terminal Help

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

main.py New Request

Query Headers³ Auth Body Tests Pre Run

HTTP Headers

Accept */*

User-Agent Thunder Client (https://www.thunderclient.com)

X-API-Key testapikey

header value

Status: 200 OK Size: 6.52 KB Time: 10 ms

Response Headers⁵ Cookies Results Docs

```
1 {  
2   "page": 1,  
3   "page_size": 30,  
4   "total": 4,  
5   "results": [  
6     {  
7       "id": "a-feast-for-crows-a-song-of-ice-and-fire-4_357",  
8       "title": "A Feast for Crows (A Song of Ice and Fire #4)",  
9       "description": "With A Feast for Crows, Martin delivers the long-awaited fourth volume of the landmark series that has redefined imaginative fiction and stands as a modern masterpiece in the making. After centuries of bitter strife, the seven powers dividing the land have beaten one another into an uneasy truce. But it's not long before the survivors, outlaws, renegades, and carrion eaters With A Feast for Crows, Martin delivers the long-awaited fourth volume of the landmark series that has redefined imaginative fiction and stands as a modern masterpiece in the making. After centuries of bitter strife, the seven powers dividing the land have beaten one another into an uneasy truce. But it's not long before the survivors, outlaws, renegades, and carrion eaters of the Seven Kingdoms gather. Now, as the human crow assemble over a banquet of ashes, daring new plots and dangerous new alliances are formed while surprising faces—some familiar, others only just appearing—emerge from an ominous twilight of past struggles and chaos to take up the challenges of the terrible times ahead. Nobles and commoners, soldiers and sorcerers, assassins and sages, are coming together to stake their fortunes... and their lives. For at a feast for crows, many are the guests—but only a few are the survivors. ...more",  
10      "category": "Fantasy",  
11    }  
12  ]  
13}  
14
```

Response Chart

11:18 PM 11/29/2025

14) http://127.0.0.1:8000/books/its-only-the-himalayas_981

THUNDER CLIENT

File Edit Selection View Go Run Terminal Help

New Request

Activity Collections Env

filter activity

GET localhost:8000/books just now

main.py New Request

Query Headers³ Auth Body Tests Pre Run

HTTP Headers

Accept */*

User-Agent Thunder Client (https://www.thunderclient.com)

X-API-Key testapikey

header value

Status: 200 OK Size: 1.85 KB Time: 6 ms

Response Headers⁵ Cookies Results Docs

```
1 {  
2   "id": "its-only-the-himalayas_981",  
3   "title": "It's Only the Himalayas",  
4   "description": "Wherever you go, whatever you do, just . . . don't do anything stupid." -My MotherDuring her yearlong adventure backpacking from South Africa to Singapore, S. Bedford definitely did a few things her mother might classify as "stupid." She swam with great white sharks in South Africa, ran from lions in Zimbabwe, climbed a Himalayan mountain without training in Nepal, and wa "wherever you go, whatever you do, just . . . don't do anything stupid." -My MotherDuring her yearlong adventure backpacking from South Africa to Singapore, S. Bedford definitely did a few things her mother might classify as "stupid." She swam with great white sharks in South Africa, ran from lions in Zimbabwe, climbed a Himalayan mountain without training in Nepal, and watched as her friend was attacked by a monkey in Indonesia .But interspersed in those slightly more crazy moments, Sue Bedford and her friend "Sara the Stoic" experienced the sights, sounds, life, and culture of fifteen countries. Joined along the way by a few friends and their aging fathers here and there, Sue and Sara experience the trip of a lifetime. They fall in love with the world, cultivate an appreciation for home, and discover who, or what, they want to become. It's Only the Himalayas is the incredibly funny, sometimes outlandish, always entertaining confession of a young backpacker that will inspire you to take your own adventure. ...more",  
5   "category": "Travel",  
6   "price_including_tax": 45.17,  
7   "price_excluding_tax": 45.17,  
8   "availability": "In stock (19 available)",  
9   "num_reviews": 0,  
10}  
11
```

Response Chart

11:21 PM 11/29/2025

15) http://127.0.0.1:8000/changes (after Scheduler crawl)

The screenshot shows the Web Crawler interface in VS Code. A GET request is made to `http://127.0.0.1:8000/changes`. The response body is a JSON object with one line of code:

```
1 {  
2     "results": []  
3 }
```

16) http://127.0.0.1:8000/changes (after 1st crawl of the website)

The screenshot shows the Web Crawler interface in VS Code. A GET request is made to `http://127.0.0.1:8000/changes`. The response body is a JSON object with three lines of code, representing the results of the first crawl:

```
1 {  
2     "results": [  
3         {  
4             "_id": "692e8d207da91191e2f92263",  
5             "book_id": "bleach-vol-1-strawberry-and-the-soul-reapers-bleach-1_7",  
6             "change_type": "new",  
7             "recent": "new",  
8             "timestamp": "2025-12-02T06:54:24.399834+00:00Z"  
9         },  
10         {  
11             "_id": "692e8d207da91191e2f92261",  
12             "book_id": "a-spy-s-devotion-the-regency-spies-of-london-1_3",  
13             "change_type": "new",  
14             "recent": "new",  
15             "timestamp": "2025-12-02T06:54:24.034272+00:00Z"  
16         },  
17         {  
18             "_id": "692e8d1f7da91191e2f9225f",  
19             "book_id": "1000-places-to-see-before-you-die_1",  
20         }  
21     ]  
22 }
```