

## Test Plan

### 1.Prerequisite: Unit Testing (Python)

### 2.Envirment setup and configuration:

- A. software: VS code
- B. operating system:MacOs
- C. hardware:laptop

3.

result of pylint

```
(base) Rongs-MacBook-Pro:sp21-cs242-assignment2 rongni$ pylint scraper.py
***** Module scraper
scraper.py:12:0: W0212: Access to a protected member _create_default_https_context of a client class (protected-access)
scraper.py:12:36: W0212: Access to a protected member _create_unverified_context of a client class (protected-access)
scraper.py:48:23: W1202: Use lazy % formatting in logging functions (logging-format-interpolation)
scraper.py:50:8: W0150: return statement in finally block may swallow exception (lost-exception)
scraper.py:4:0: C0411: third party import "import pymongo" should be placed before "from json_convert import write_json" (wrong-import-order)
scraper.py:5:0: C0411: standard import "import time" should be placed before "from dotenv import load_dotenv" (wrong-import-order)
scraper.py:6:0: C0411: standard import "import argparse" should be placed before "from dotenv import load_dotenv" (wrong-import-order)
scraper.py:7:0: C0411: standard import "import os" should be placed before "from dotenv import load_dotenv" (wrong-import-order)
scraper.py:8:0: C0411: standard import "import ssl" should be placed before "from dotenv import load_dotenv" (wrong-import-order)
scraper.py:9:0: C0411: standard import "from urllib.request import urlopen" should be placed before "from dotenv import load_dotenv" (wrong-import-order)
scraper.py:10:0: C0411: third party import "from bs4 import BeautifulSoup as soup" should be placed before "from json_convert import write_json" (wrong-import-order)
scraper.py:11:0: C0411: standard import "import logging" should be placed before "from dotenv import load_dotenv" (wrong-import-order)

Your code has been rated at 9.21/10 (previous run: 10.00/10, -0.79)
```

```
(base) Rongs-MacBook-Pro:sp21-cs242-assignment2 rongni$ pylint json_convert.py
```

```
-----
Your code has been rated at 10.00/10 (previous run: 10.00/10, +0.00)
```

#### 4. test scraper.py

```
parser.add_argument('--num_book', type=int, default=200)
parser.add_argument('--num_author', type=int, default=50)
parser.add_argument(
    '--url', type=str, default="https://www.goodreads.com/book/show/149267.The_Stand")
parser.add_argument('--export_to_json', type=bool, default=True)
args = parser.parse_args()
```

the command line interface will accept up to four inputs number of books, number of authors, starting page url, and boolean option for export the scraped data as a json file .

If the user put the invalid url, the log file will have output show below

```
(base) Rongs-MacBook-Pro:sp21-cs242-assignment2 rongni$ python3 scraper.py --url dsdsds
```

```
2021-03-01 20:20:16,050 Unknown url(dsdsds)
```

if the user input of number of books and num of author is not integer will got error like below

```
(base) Rongs-MacBook-Pro:sp21-cs242-assignment2 rongni$ python3 scraper.py --num_book a
usage: scraper.py [-h] [--num_book NUM_BOOK] [--num_author NUM_AUTHOR] [--url URL] [--export_to_json EXPORT_TO_JSON]
scraper.py: error: argument --num_book: invalid int value: 'a'
```

the scraper.log will also show the progress of scraping

```
2021-03-01 20:22:51,939 getting Books
2021-03-01 20:22:57,341 getting author
|
```

the scraped data will directly insert to the mongoDB database. I used the cloud version(atlas) to store all data shown below.

### goodreaddb.author

COLLECTION SIZE: 436.02KB TOTAL DOCUMENTS: 295 INDEXES TOTAL SIZE: 44KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

FILTER {"filter":"example"}

QUERY RESULTS 1-20 OF MANY

```
_id: ObjectId("603ca5c9fb6ef196de48d384")
name: "George Orwell"
author_url: "https://www.goodreads.com/author/show/3706.George_Orwell"
author_id: "3706"
rating: "4.09"
rating_count: "6646269"
review_count: "154198"
image_url: "https://images.gr-assets.com/authors/1588856560p8/3706.jpg"
> related_authors: Array
> author_books: Array
```

```
>
_id: ObjectId("603ca5d2fb6ef196de48d721")
name: "J.D. Salinger"
author_url: "https://www.goodreads.com/author/show/819789.J_D_Salinger"
author_id: "819789"
rating: "3.84"
rating_count: "3232472"
review_count: "74005"
image_url: "https://images.gr-assets.com/authors/1288777679p8/819789.jpg"
> related_authors: Array
> author_books: Array
```

```
_id: ObjectId("603ca5e4fb6ef196de48ddaf")
name: "Jane Austen"
author_url: "https://www.goodreads.com/author/show/1265.Jane_Austen"
author_id: "1265"
rating: "4.13"
rating_count: "6320529"
review_count: "183445"
image_url: "https://images.gr-assets.com/authors/1588941810p8/1265.jpg"
> related_authors: Array
> author_books: Array
```



# goodreaddb.book

COLLECTION SIZE: 1.07MB    TOTAL DOCUMENTS: 627    INDEXES TOTAL SIZE: 52KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes ●

FILTER {"filter":"example"}

## QUERY RESULTS 1-20 OF MANY

```
_id: ObjectId("603ca5b6fb6ef196de48ce0a")
book_url: "https://www.goodreads.com/book/show/2657.To_Kill_a_Mockingbird"
title: "To Kill a Mockingbird"
book_id: "2657"
ISBN: ""
author_url: "https://www.goodreads.com/author/show/1825.Harper_Lee"
author: "Harper Lee"
rating: "4.28"
rating_count: "4702893"
review_count: "94158"
image_url: "https://i.gr-assets.com/images/S/compressed.photo.goodreads.com/books/..."
> similar_books: Array
```

>

```
_id: ObjectId("603ca5c0fb6ef196de48d149")
book_url: "https://www.goodreads.com/book/show/4671.The_Great_Gatsby"
title: "The Great Gatsby"
book_id: "4671"
ISBN: ""
author_url: "https://www.goodreads.com/author/show/3190.F_Scott_Fitzgerald"
author: "F. Scott Fitzgerald"
rating: "3.93"
rating_count: "3937421"
review_count: "70393"
image_url: "https://i.gr-assets.com/images/S/compressed.photo.goodreads.com/books/..."
> similar_books: Array
```

```
_id: ObjectId("603ca5c9fb6ef196de48d390")
book_url: "https://www.goodreads.com/book/show/40961427-1984"
title: "1984"
book_id: "409614271984"
ISBN: ""
author_url: "https://www.goodreads.com/author/show/3706.George_Orwell"
author: "George Orwell"
rating: "4.19"
rating_count: "3319849"
review_count: "74526"
image_url: "https://i.gr-assets.com/images/S/compressed.photo.goodreads.com/books/..."
> similar_books: Array
```