

Manual Test Plan for sp21-CS242-assignment 2.0

Table of Contents

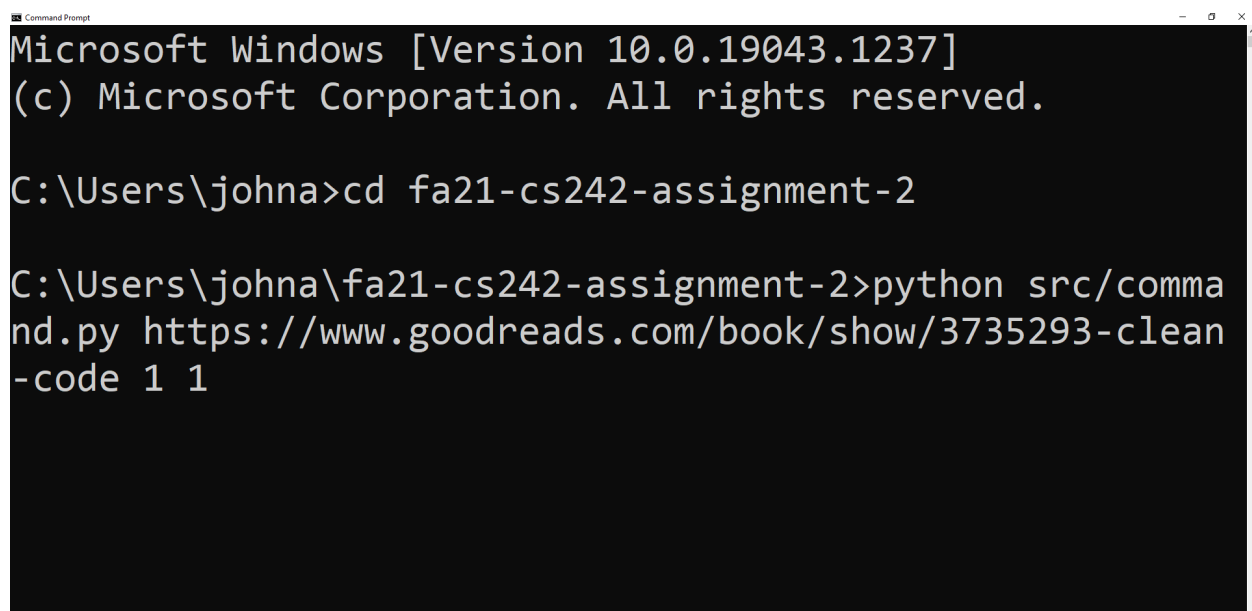
- Environment Setup
- Test - User Input in Command Line Interface
- Test - Scraping/Database Handling

Environment Setup

- Python 3.0
- unittest
- Eclipse - 4.20.0
- Windows 10

Test - User Input in Command Line Interface

Test-1.1:



```
Microsoft Windows [Version 10.0.19043.1237]
(c) Microsoft Corporation. All rights reserved.

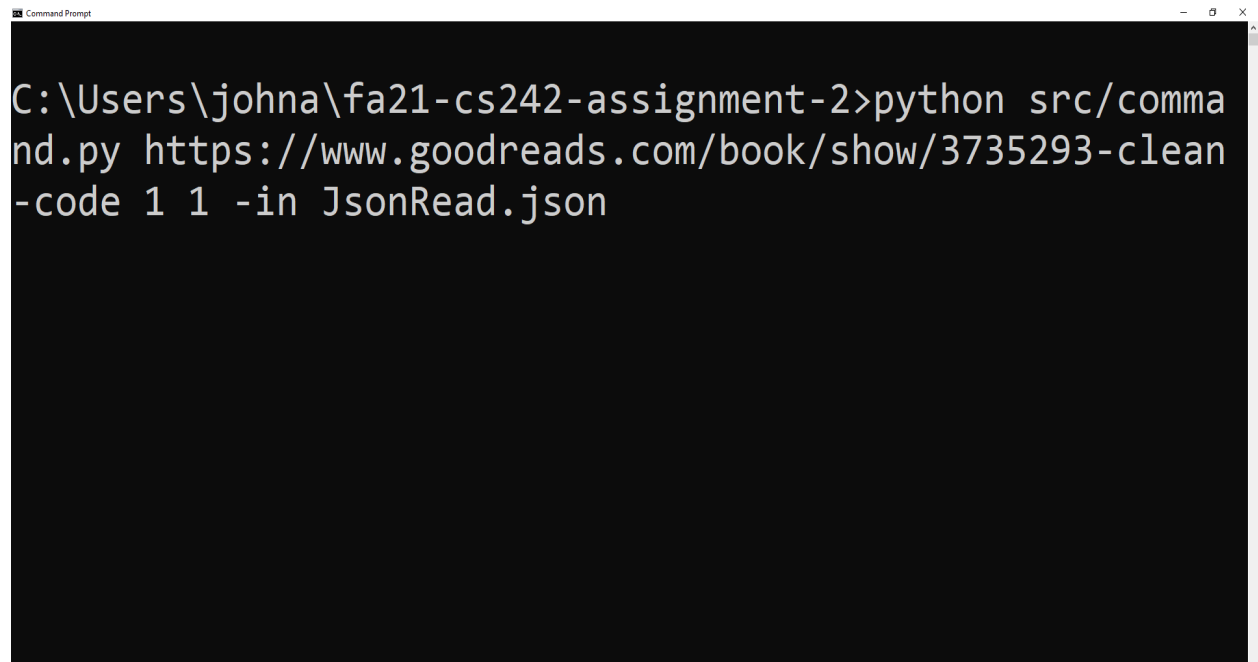
C:\Users\johna>cd fa21-cs242-assignment-2

C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1
```

@Test 1.1 Basic Command Line Arguments

- The user must input a minimum of a starting book link, the number of books, and the number of authors to scrape.

Test-1.2:

A screenshot of a Windows Command Prompt window. The title bar at the top reads "Command Prompt". The command prompt shows the following text:

```
C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1 -in JsonRead.json
```

The command is being entered on a single line. The prompt character is a greater-than sign (>). The command itself is `python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1 -in JsonRead.json`. The command prompt window has a standard Windows interface with a title bar, a menu bar (File, Edit, Format, View, Tools, Help), and a scroll bar on the right.

@Test 1.2 Optional Command Line Argument - JSON File to Read

- The user can input a valid JSON File to read from along with the basic input requirements and send data to the MongoDB database.

Test-1.3:

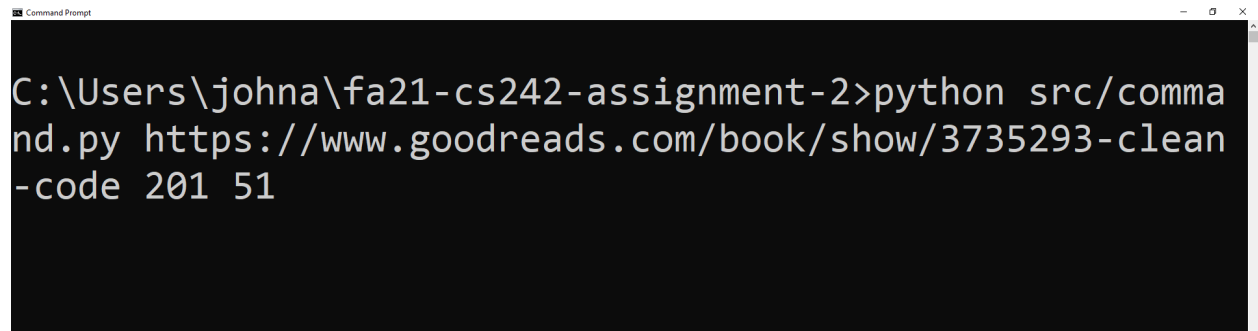
A screenshot of a Windows Command Prompt window. The title bar at the top reads "Command Prompt". The command prompt shows the following command being entered: `C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1 -existing https://www.goodreads.com/book/show/3735293-clean-code -out output.json`

```
C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1 -existing https://www.goodreads.com/book/show/3735293-clean-code -out output.json
```

@Test 1.3 Optional Command Line Argument - JSON File Export and Existing URL

- The user can also input the optional arguments of an output file of their choice, along with a valid link to the GoodReads website that can be either a book or an author. The data from the database will be exported to the output file if these arguments are supplied.

Test 1.4 - Warning For High Input of Books or Authors

A screenshot of a Windows Command Prompt window. The title bar at the top reads "Command Prompt". The command prompt shows the following text:

```
C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 201 51
```

```
C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 201 51
```

@Test 1-4

- If a user inputs an amount of books or authors greater than 200 and 50 respectively, the program will print a warning that states that this could be harmful to the website being scraped.

Test 1.5 - Warning For Links Contained within Robots.txt

```
# See http://www.robotstxt.org/robotstxt.html for documentation on how to use the robots.txt file
User-agent: *
Disallow: /about/team_member/
Disallow: /admin
Disallow: /api
Disallow: /blog/list_rss
Disallow: /book/reviews/
Disallow: /book_link/follow/
Disallow: /buy_buttons/
Disallow: /ebooks
Disallow: /event/show/
Disallow: /home/index_rss
Disallow: /oggiPlayerLoader.htm
Disallow: /photo/group/
Disallow: /quotes/list_rss
Disallow: /reader
Disallow: /review/list_rss
Disallow: /review/rate
Disallow: /shelf/user_shelves
Disallow: /story
Disallow: /tooltips
Disallow: /track
Disallow: /trivia/answer
Disallow: /user/updates_rss
Disallow: /book/show/52976360
Disallow: /book/show/53024603
Disallow: /book/show/52667524
Disallow: /book/show/13634645
Disallow: /book/show/53203379
Disallow: /book/show/53472095
Disallow: /book/show/53249757
Disallow: /book/show/50692110
Disallow: /book/show/53451162
Disallow: /book/show/53005407
Disallow: /book/show/50990510
Disallow: /book/show/52558212
Disallow: /book/show/52799562
Disallow: /book/show/53179891
Disallow: /book/show/53245386
Disallow: /book/show/52719017
Disallow: /book/show/53085988
Disallow: /book/show/51038483
Disallow: /book/show/51329187
Disallow: /book/show/53280099
Disallow: /book/show/52717468
Disallow: /book/show/52976361
Disallow: /book/show/53472951
Disallow: /book/show/51108563
Disallow: /book/show/53050699
Disallow: /book/show/51918376
Disallow: /book/show/51434342
Disallow: /book/show/53201597
Disallow: /book/show/53172541
```

@Test 1-5

- If a user inputs a link that is contained within the Robots.txt file of GoodReads.com, the program will print a warning message. Also, during traversal the program will skip over any urls that are within the Robots.txt file.

Test : Scraping/Database Handling

Test 1.1 Basic Program Run

```
Command Prompt

C:\Users\johna\fa21-cs242-assignment-2>python src/command.py https://www.goodreads.com/book/show/3735293-clean-code 1 1
Begin scraping a page
Ended this scrape
Begin scraping a page
Ended this scrape
```

+ Create Database

Q NAMESPACES

▼ GoodReadData

Authors

Books

GoodReadData.Books

COLLECTION SIZE: 2.59KB TOTAL DOCUMENTS: 1 INDEXES TOTAL SIZE: 20KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

INSERT DOCUMENT

FILTER { field: 'value' } OPTIONS Apply

QUERY RESULTS 1-1 OF 1

```
_id: ObjectId("615125895c60a0ac3c0bd239")
book_url: "https://www.goodreads.com/book/show/3735293-clean-code"
title: "Clean Code: A Handbook of Agile Software Craftsmanship"
book_id: "3735293"
ISBN: "9780132350884"
author_url: "https://www.goodreads.com/author/show/45372.Robert_C_Martin"
author: "Robert C. Martin"
rating: "4.40"
rating_count: "17128"
review_count: "1031"
image_url: "https://i.gr-assets.com/images/S/compressed.photo.goodreads.com/books/..."
similar_books: Array
```

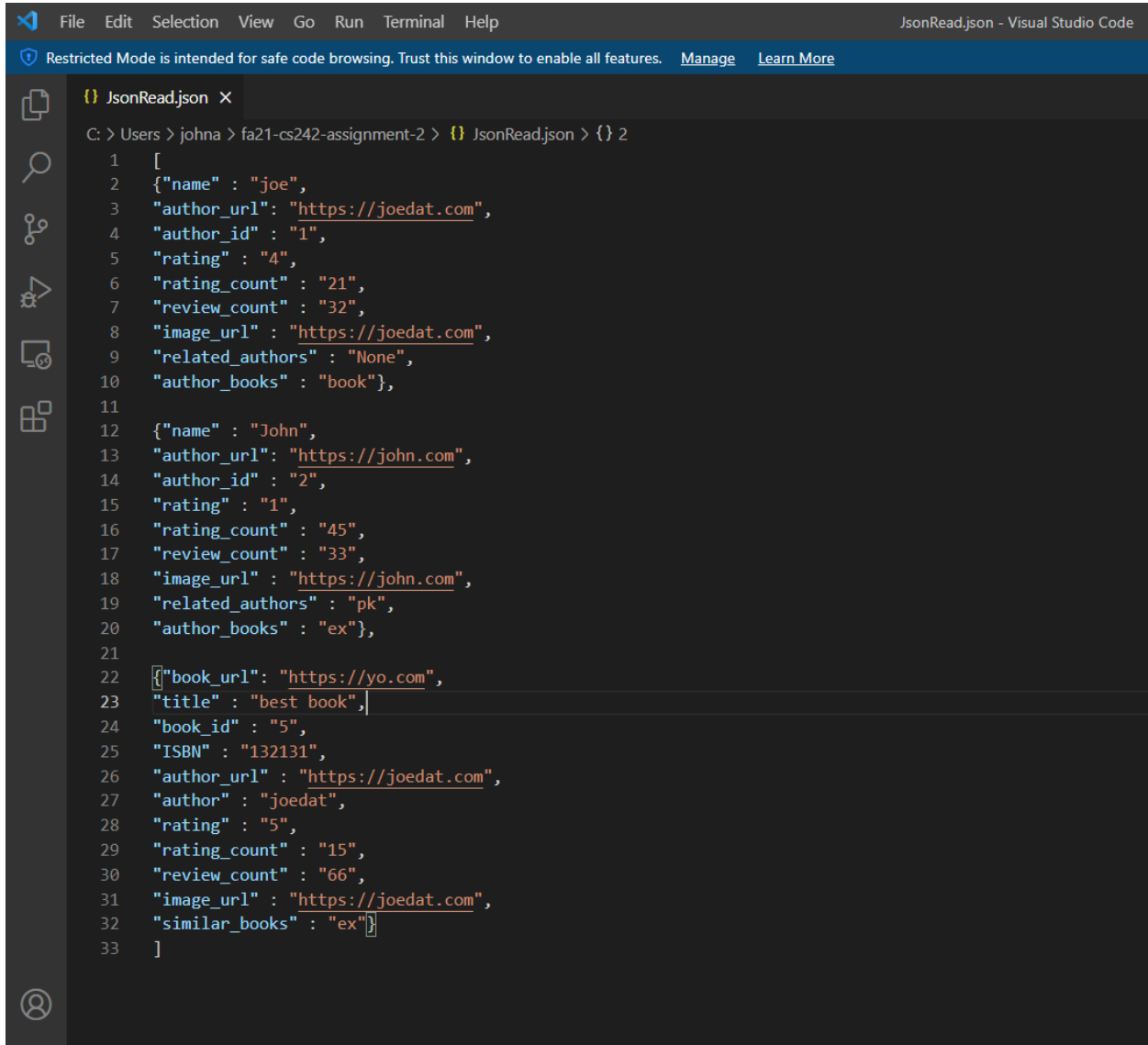
The screenshot shows a MongoDB database interface. On the left, a sidebar lists the database structure: 'GoodReadData' (expanded) contains 'Authors' and 'Books'. The main panel is titled 'GoodReadData.Authors' and shows collection statistics: 'COLLECTION SIZE: 1.83KB', 'TOTAL DOCUMENTS: 1', and 'INDEXES TOTAL SIZE: 4KB'. The 'Find' tab is active, displaying a filter bar with the query '{ field: 'value' }'. Below the filter bar, the query results are shown as a single document for Robert C. Martin.

```
{
  "_id": ObjectId("615125b85c60a0ac3c0bf005"),
  "name": "Robert C. Martin",
  "author_url": "https://www.goodreads.com/author/show/45372.Robert_C_Martin",
  "author_id": "45372",
  "rating": "4.34",
  "rating_count": "31523",
  "review_count": "2094",
  "image_url": "https://images.gr-assets.com/authors/1490470967p5/45372.jpg",
  "related_authors": Array,
  "author_books": Array
}
```

@Test 1.1

- If a user inputs the 3 basic arguments, the program will proceed to handle scraping and transferring of data to the database in that order. The database data is transferred after each unique scrape is completed. Messages are printed to let the user know when each scrape has started or ended. In this case, the user wanted to scrape and transfer 1 author and 1 book, which is in accordance with the contents of the database.

Test 1.2: Optional Argument JSON File to Be Read



```
{} JsonRead.json X
C: > Users > johna > fa21-cs242-assignment-2 > {} JsonRead.json > {} 2
1  [
2  {"name" : "joe",
3   "author_url": "https://joedat.com",
4   "author_id" : "1",
5   "rating" : "4",
6   "rating_count" : "21",
7   "review_count" : "32",
8   "image_url" : "https://joedat.com",
9   "related_authors" : "None",
10  "author_books" : "book"},
11
12  {"name" : "John",
13   "author_url": "https://john.com",
14   "author_id" : "2",
15   "rating" : "1",
16   "rating_count" : "45",
17   "review_count" : "33",
18   "image_url" : "https://john.com",
19   "related_authors" : "pk",
20   "author_books" : "ex"},
21
22  [{"book_url": "https://yo.com",
23   "title" : "best book",
24   "book_id" : "5",
25   "ISBN" : "132131",
26   "author_url" : "https://joedat.com",
27   "author" : "joedat",
28   "rating" : "5",
29   "rating_count" : "15",
30   "review_count" : "66",
31   "image_url" : "https://joedat.com",
32   "similar_books" : "ex"}]
33  ]
```


QUERY RESULTS 1-3 OF 3

```
>
  _id: ObjectId("615125b85c60a0ac3c0bf005")
  name: "Robert C. Martin"
  author_url: "https://www.goodreads.com/author/show/45372.Robert_C_Martin"
  author_id: "45372"
  rating: "4.34"
  rating_count: "31523"
  review_count: "2094"
  image_url: "https://images.gr-assets.com/authors/1490470967p5/45372.jpg"
  > related_authors: Array
  > author_books: Array
```

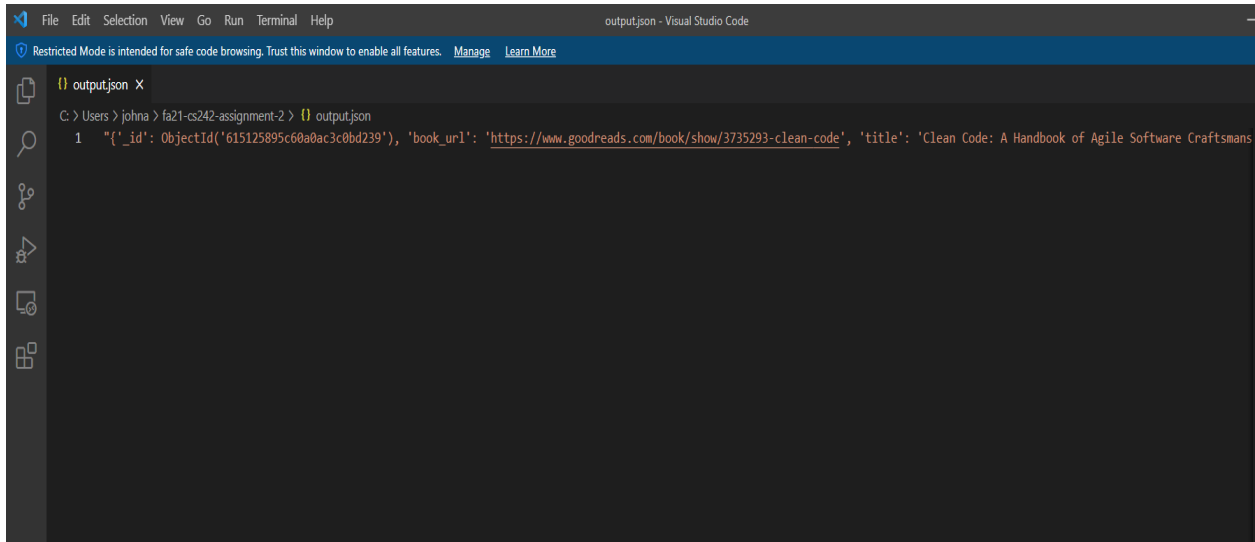
```
  _id: ObjectId("615131375c60a0ac3c135b75")
  name: "joe"
  author_url: "https://joedat.com"
  author_id: "1"
  rating: "4"
  rating_count: "21"
  review_count: "32"
  image_url: "https://joedat.com"
  related_authors: "None"
  author_books: "book"
```

```
  _id: ObjectId("6151313b5c60a0ac3c135e98")
  name: "John"
  author_url: "https://john.com"
  author_id: "2"
  rating: "1"
  rating_count: "45"
  review_count: "33"
  image_url: "https://john.com"
  related_authors: "pk"
  author_books: "ex"
```

@Test 1.2

- Since the user has added the optional arguments of a JSON file to read from and update the database, the program will read the JSON file given and either create a new entry or update an existing entry in the Database. The database shows that these changes took place when supplied with the extra arguments.

Test 1.3 - Optional Arguments JSON Output File and Existing Data Entry



The screenshot shows the Visual Studio Code interface. The title bar at the top reads "output.json - Visual Studio Code". Below the title bar, a blue status bar indicates "Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More". The main editor area displays a file named "output.json" with a single line of JSON data:

```
1  {"_id": ObjectId('615125895c60a0ac3c0bd239'), 'book_url': 'https://www.goodreads.com/book/show/3735293-clean-code', 'title': 'Clean Code: A Handbook of Agile Software Craftsmans'}
```

@Test 1.3

- Since the optional arguments of a JSON output file and link to an existing book or author was supplied, the program will now also read data from the MongoDB database and send the existing entry to a specified output file, which in this case was output.json.