Manual Test Plan For Web Scraping Project

1. Pre-Test

The application is run on a Python 3.8 environment, with Pymongo 3.7.2 as connection to MongoDB database.

2. Test Plan For Scraper

A. Normal Run Command

```
(base) JoeldeMacBook-Pro:sp21-cs242-assignment2 joelzhang$ python main.py -a 2 -b 2 -i "www.website.com" -j jsonfile.json
```

B. The website is not pointing to www.goodreads.com

```
The website is not pointing to Goodreads
```

C. The website is not pointing to www.goodreads.com/book

```
The website is not pointing to a potential book page
```

D. The input argument with invalid type:

```
usage: main.py [-h] [-a AUTHOR] [-b BOOK] [-i INSERT] [-j JSON]
main.py: error: argument -a/--author: invalid int value: 'asd'
```

E. Valid scraping url with invalid json.file insertion

```
scraping book detail now...
Scraping author details now...
not able to open the file, exiting...
invalid json file
exporting from database as JSON file...
```

F. Valid input full version

```
python main.py -a 2 -b 1 -i "https://www.goodreads.com/book/show/3735293-clean-code?from_search=true&qid=HhMDV0vMa5&rank=1" -j "TestJSON/test1.json"
```

The result with progress report should be:

```
scraping book detail now...
Scraping author details now...
inserting author details from json file...
inserting author details from json file...
inserting book details from json file...
inserting book details from json file...
exporting from database as JSON file...
```

G. Connection failed

```
There is an issue on network, connection failed
```

H. More than 2000 books

```
book limit too big, more than 2000
```

I. More than 2000 authors

```
author limit too big, more than 2000
```

J. Warning for too many books and authors, but will keep running

```
Warning: scraping more than 200 could be very costly and take a long time
Warning: scraping more than 50 could be very costly and take a long time
```

3. Test Plan for API and query

Query Language Test

The following includes test plans for API and query:

Current Instruction of main.py argument:

```
scraping the website

optional arguments:
  -h, --help            show this help message and exit
  -a AUTHOR, --author AUTHOR
                        author limit of scraping
  -b BOOK, --book BOOK  book limit of scraping
  -i INSERT, --insert INSERT
                        insert url
  -j JSON, --json JSON  json file to insert
  -bq BOOKQUERY, --bookquery BOOKQUERY
                        book query language
  -aq AUTHORQUERY, --authorquery AUTHORQUERY
                        author query language
```

Using newly added -bq and -aq as query language function:

A. Invalid input for author query:

```
python main.py -aq "book.book_id"
```

```
    raise TypeError("This is not a author")
TypeError: This is not a author
```

B. Valid Id search:

```
python main.py -bq "book.book_id"
```


```
{'CleanCode:AHandbookofAgileSoftwareCraftsmanship': '3735293', '\nBuildingMicroservices:DesigningFine-GrainedSystems\n': '22512931', '\nAccelerate:BuildingandScalingHigh-PerformingTechnologyOrganizations\n': '35747076', '\nReleaseIt!:DesignandDeployProduction-ReadySoftware(PragmaticProgrammers)\n': '1069827', '\nFundamentalsofSoftwareArchitecture:AnEngineeringApproach\n': '44144493', '\nThePragmaticProgrammer:FromJourneymantoMaster\n': '4099', 'IamCSStudent': '67833', 'Jiaming Zhangs Book': '1234', 'Jiaming Zhangs': '12121'}
```


C. Valid complicated query search:

```
python main.py -bq "book.book_id:Not 12"
```


```
{'CleanCode:AHandbookofAgileSoftwareCraftsmanship': '3735293', '\nAccelerate:BuildingandScalingHigh-PerformingTechnologyOrganizations\n': '35747076', '\nReleaseIt!:DesignandDeployProduction-ReadySoftware(PragmaticProgrammers)\n': '1069827', '\nFundamentalsofSoftwareArchitecture:AnEngineeringApproach\n': '44144493', '\nThePragmaticProgrammer:FromJourneymantoMaster\n': '4099', 'IamCSStudent': '67833'}
```

D. Valid author query search:

```
python main.py -aq "author.author_url"
```

```
{'Jiaming Zhang': 'https://www.goodreads.com/author/show/45372.Robert_C_Martin', 'Kent Beck': 'https
://www.goodreads.com/author/show/25211.Kent_Beck', 'Cynthia Andres': 'https://www.goodreads.com/auth
or/show/38241.Cynthia_Andres', 'Sam Newman': 'https://www.goodreads.com/author/show/3362774.Sam_Newm
an', 'Nicole Forsgren': 'https://www.goodreads.com/author/show/17037914.Nicole_Forsgren', 'Jez Humbl
e': 'https://www.goodreads.com/author/show/4149510.Jez_Humble', 'Gene Kim': 'https://www.goodreads.c
om/author/show/328437.Gene_Kim', 'Michael T. Nygard': 'https://www.goodreads.com/author/show/6089.Mi
chael_T_Nygard', 'Mark   Richards': 'https://www.goodreads.com/author/show/19751098.Mark_Richards',
'Neal Ford': 'https://www.goodreads.com/author/show/715981.Neal_Ford', 'Andy Hunt': 'https://www.goo
dreads.com/author/show/2815.Andy_Hunt', 'Dave Thomas': 'https://www.goodreads.com/author/show/13.Dav
e_Thomas'}
```

E. Valid author complicated query search:

```
python main.py -aq "author.author_id:AND
```

```
{'Andy Hunt': '28151', 'Jiaming Zhang': '12345', 'Jiaming Zhangs': '1234', 'Cynthia Andres': '38241'
, 'Kent Beck': '25211'}
```

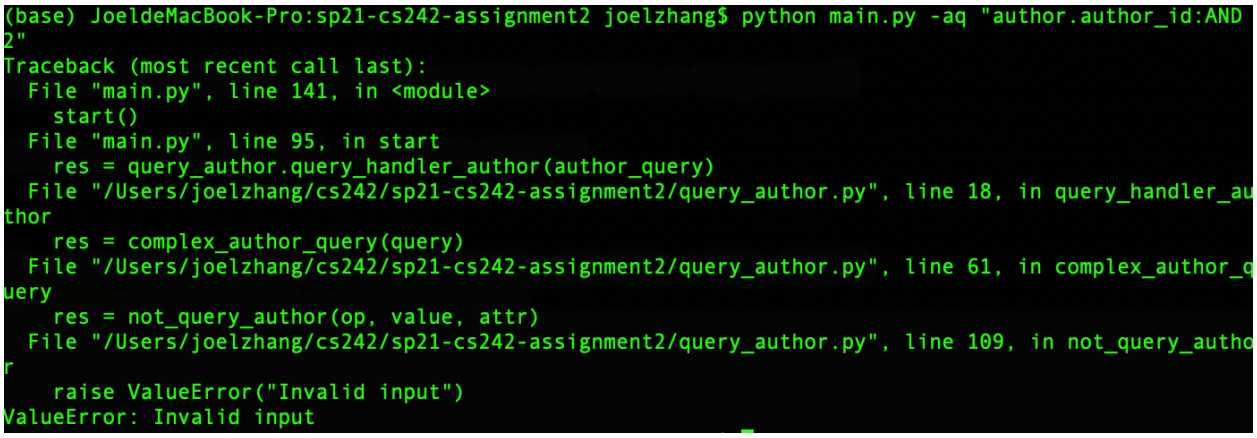
F. Invalid operator:(Not an AND or OR or NOT operator)

```
(base) JoeldeMacBook-Pro:sp21-cs242-assignment2 joelzhang$ python main.py -aq "author.author_id:AD 2
 1"
Traceback (most recent call last):
  File "main.py", line 141, in <module>
    start()
  File "main.py", line 95, in start
    res = query_author.query_handler_author(author_query)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 18, in query_handler_au
thor
    res = complex_author_query(query)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 66, in complex_author_q
uery
    res = logical_query_author(op, value_left, value_right, attr)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 130, in logical_query_a
uthor
    raise ValueError("Invalid input")
ValueError: Invalid input
```

G. Invalid number of arguments: (AND with one args)

API TEST

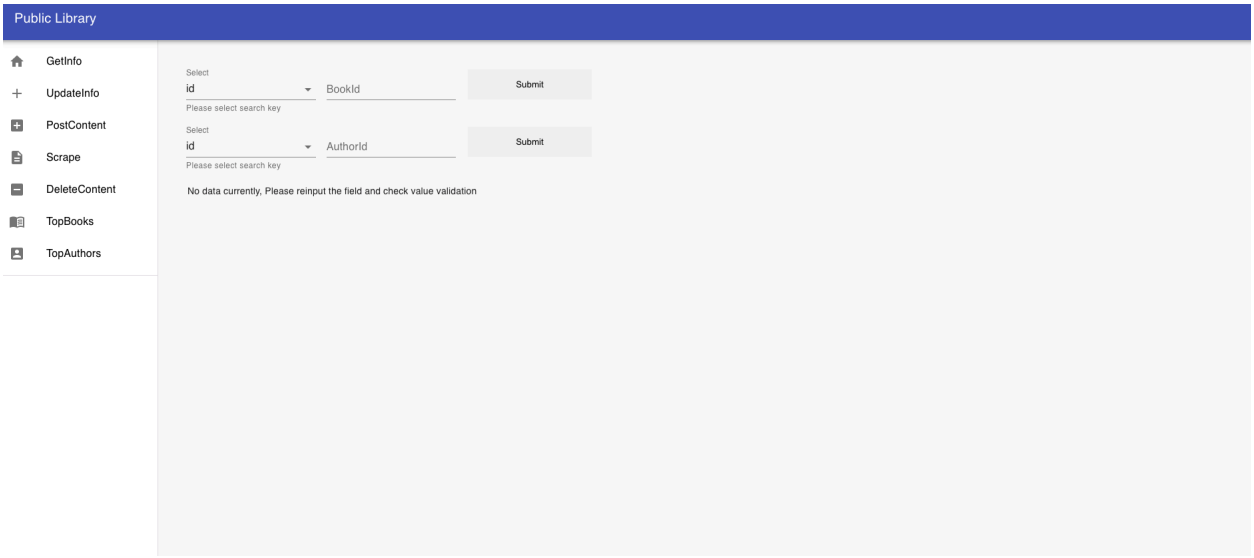
Including in Unit-test file in the folder.



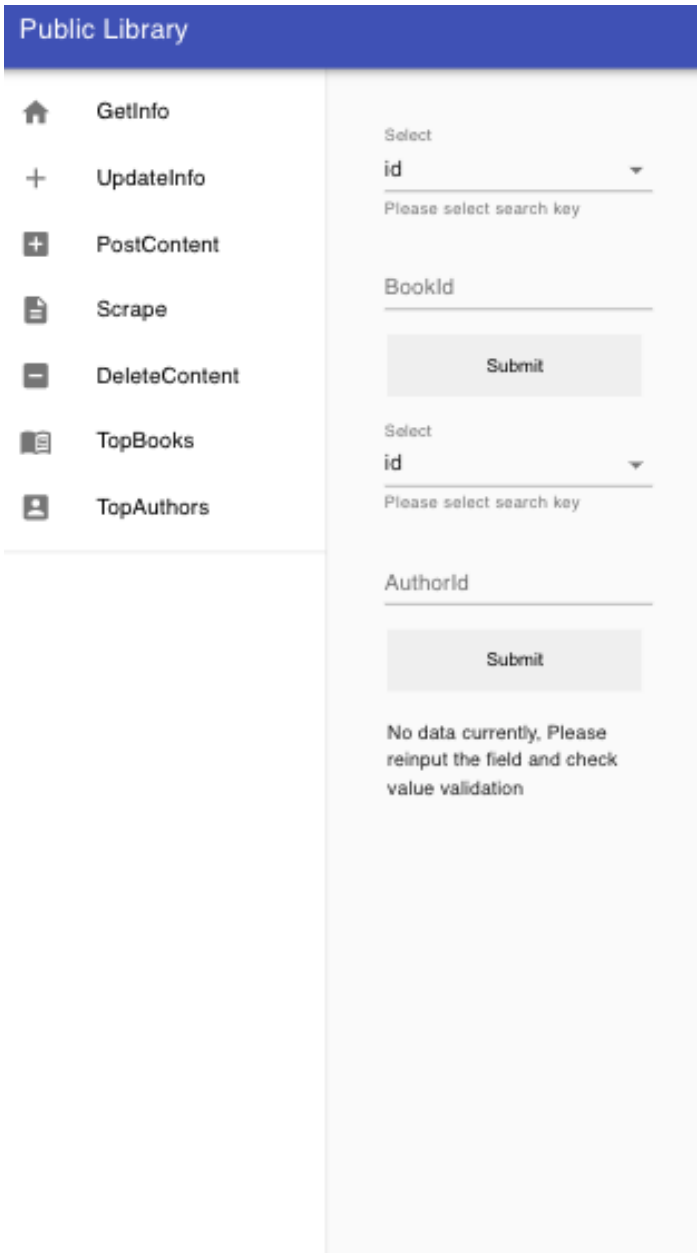
Visualization Single Web App Manual Test Plan

1. General Main Page

The page appearance should be as first opened:

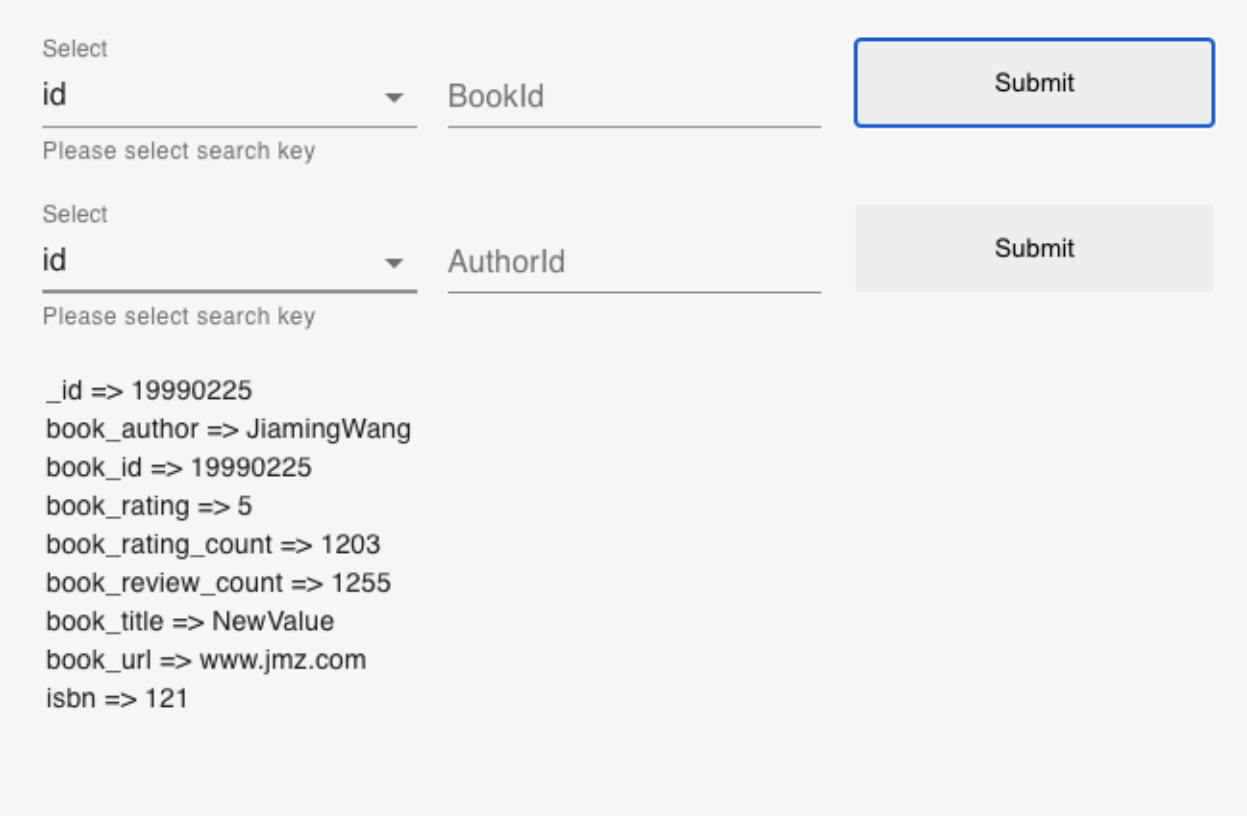


2. Responsive appearance on mobile device:

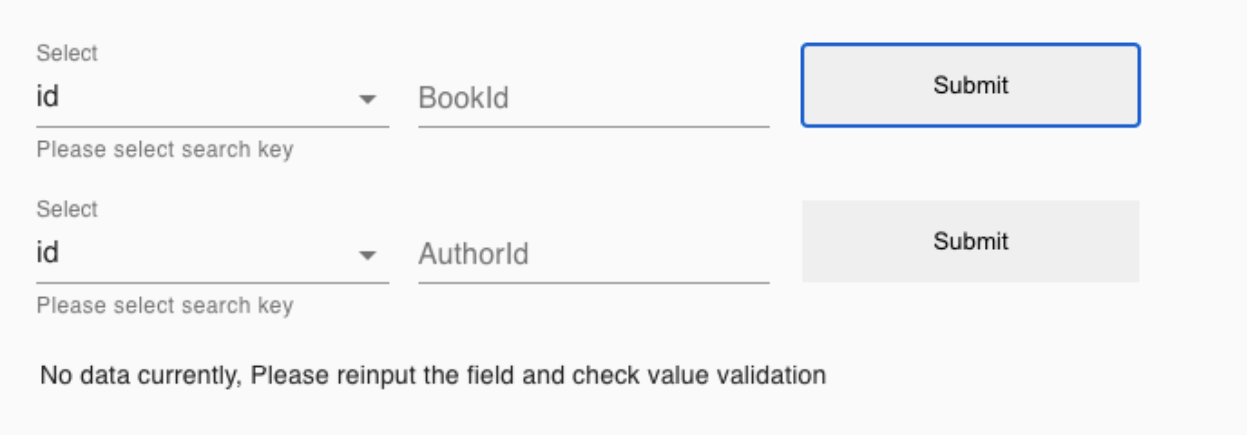


Appearance on iPhone 5 emulator.

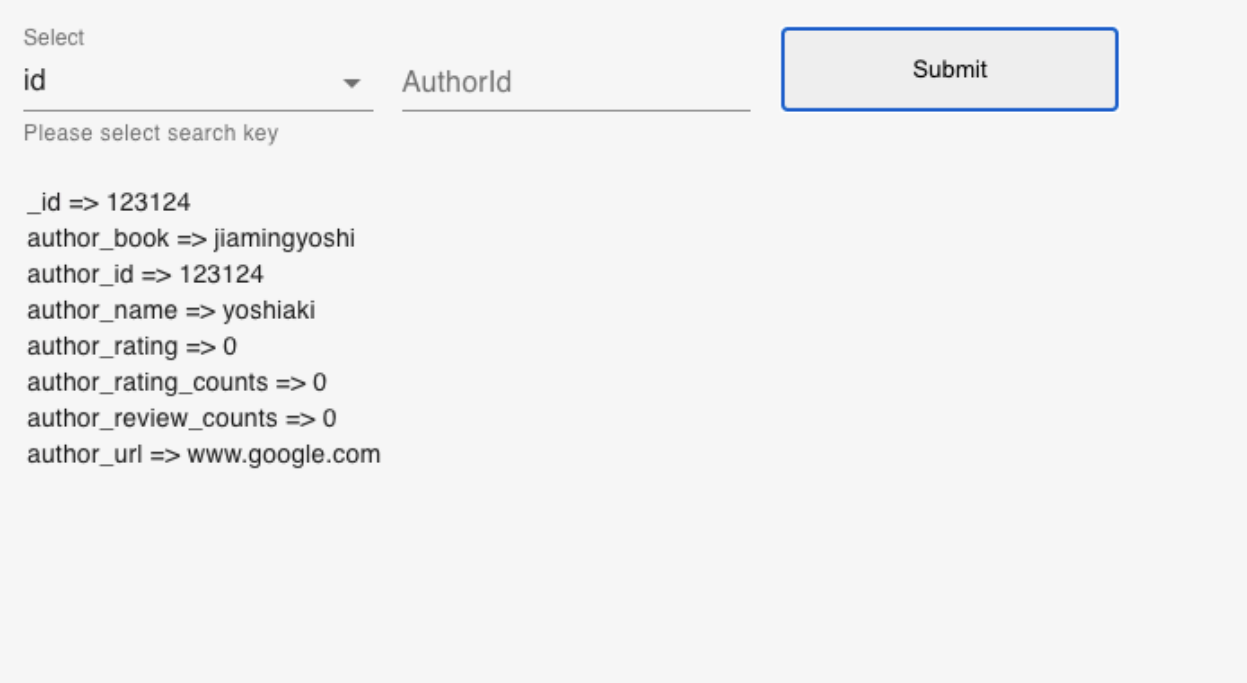
3. Successfully get book by id



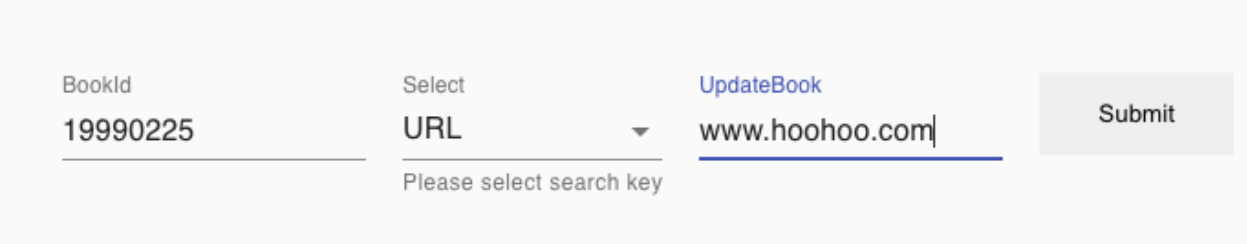
4. Failed to get book from database

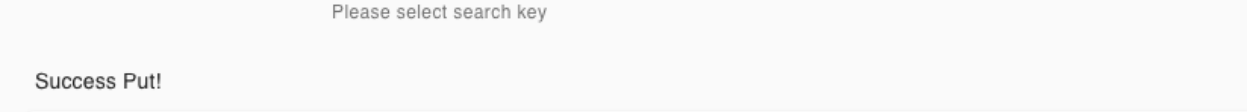


5. Successfully get author by id:

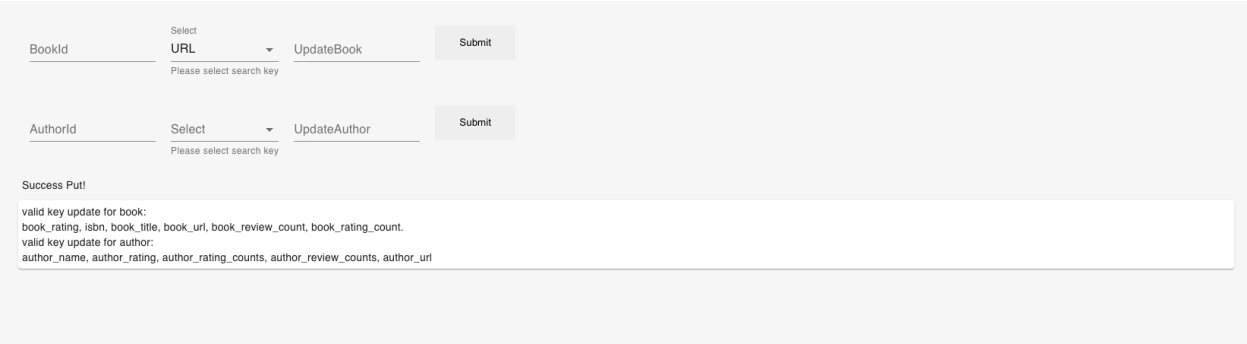


6. Update book url by id

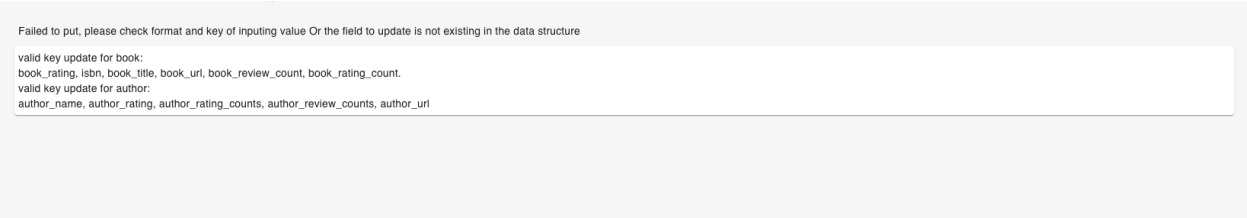




Success information output



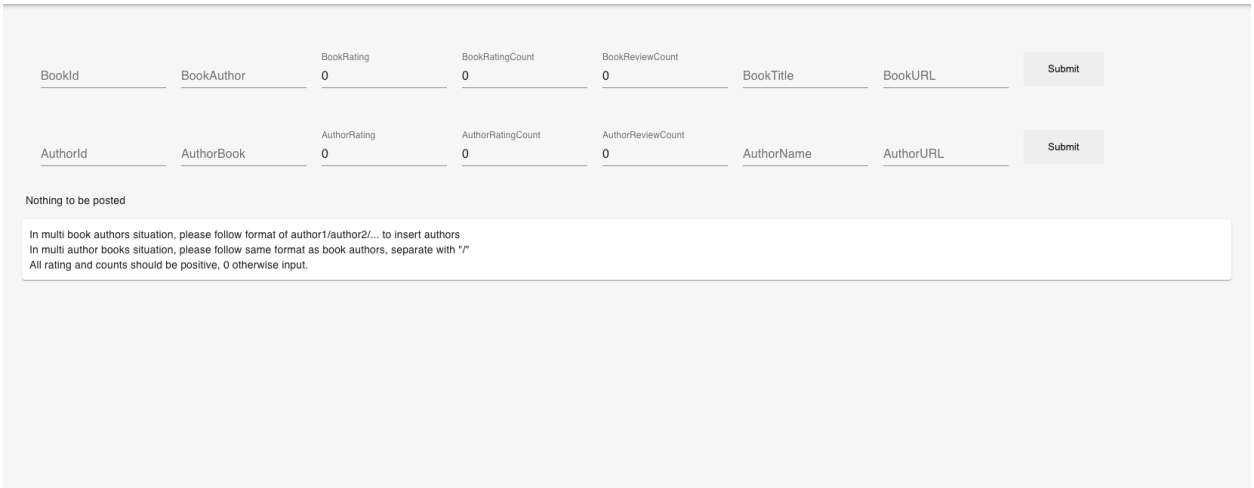
Overall page



Warning and hint information

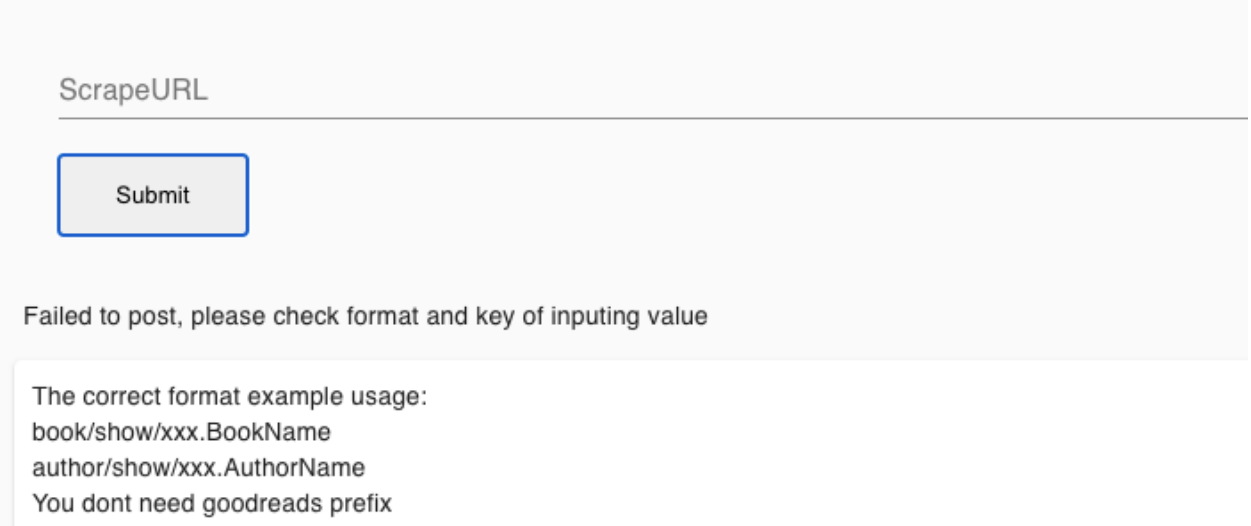
7. Post

Overall Appearance:

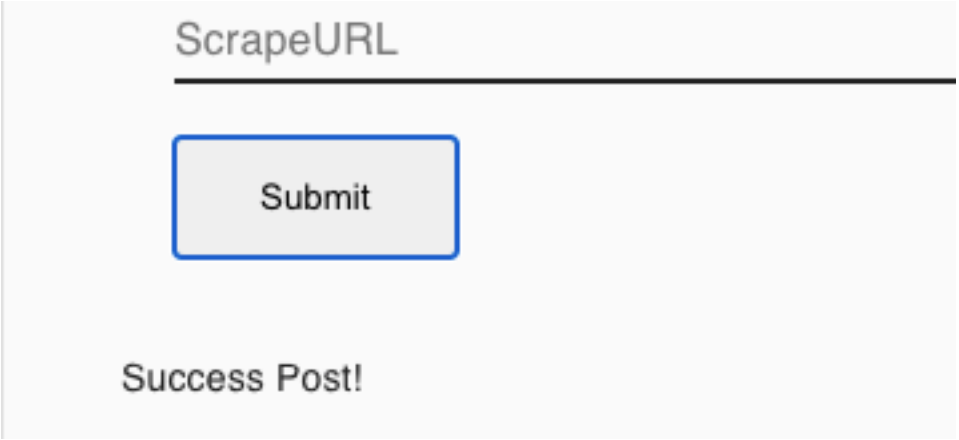


Success inform

8. Scrape and insert to database



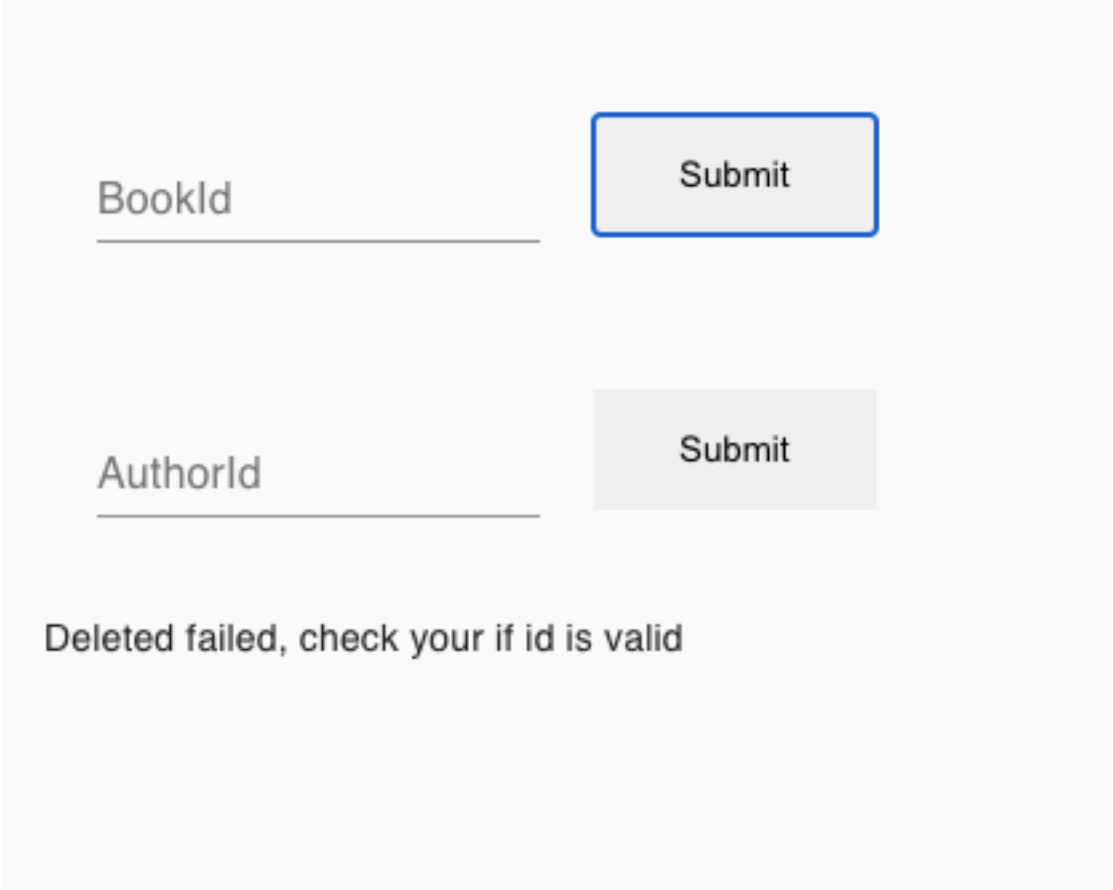
Failed to get and hint information

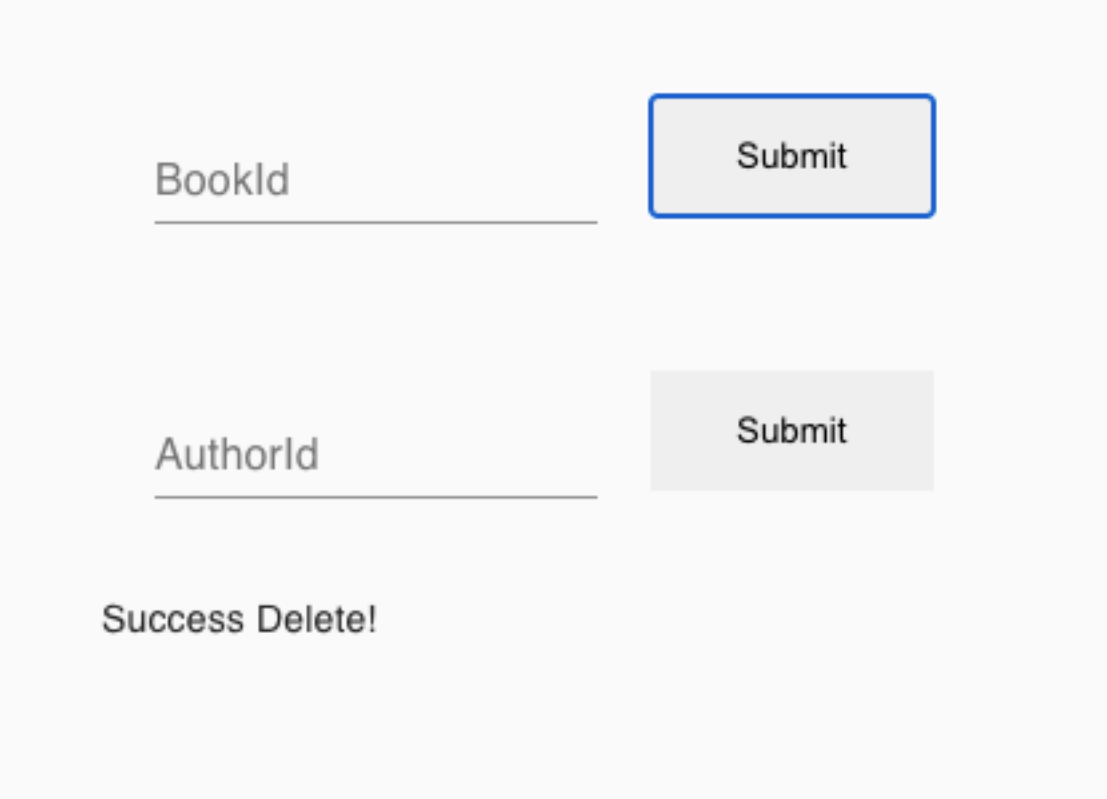


Success Posting

9. Delete a book by id

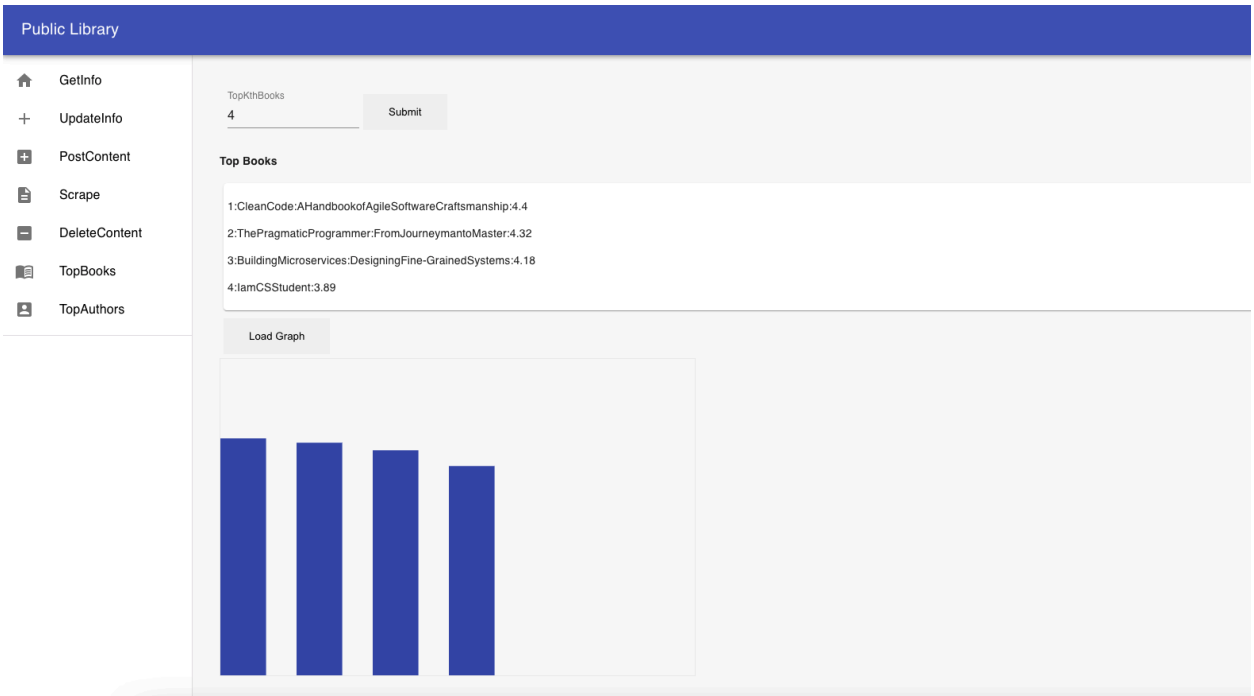
Failed





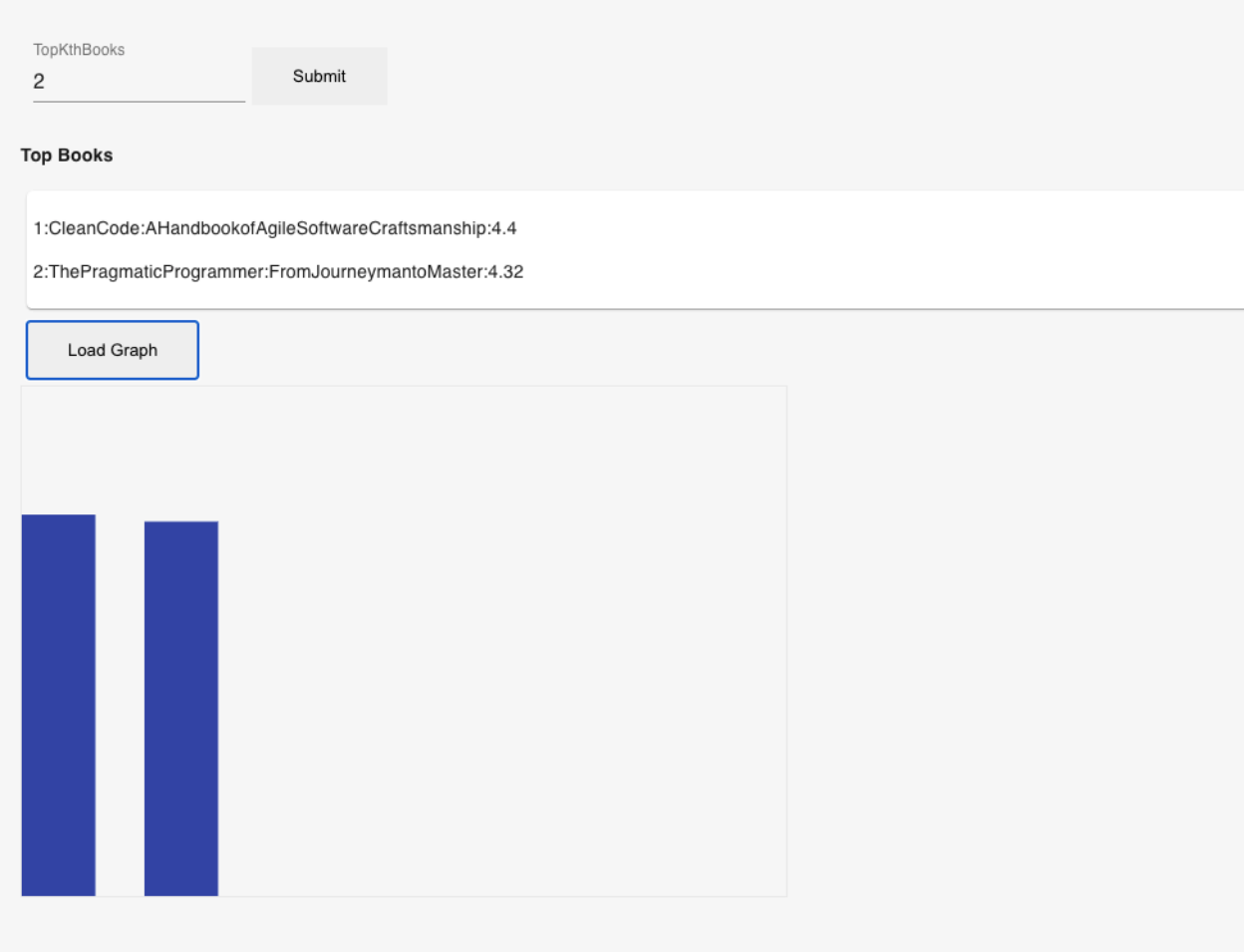
Success

9. Top Books Visual

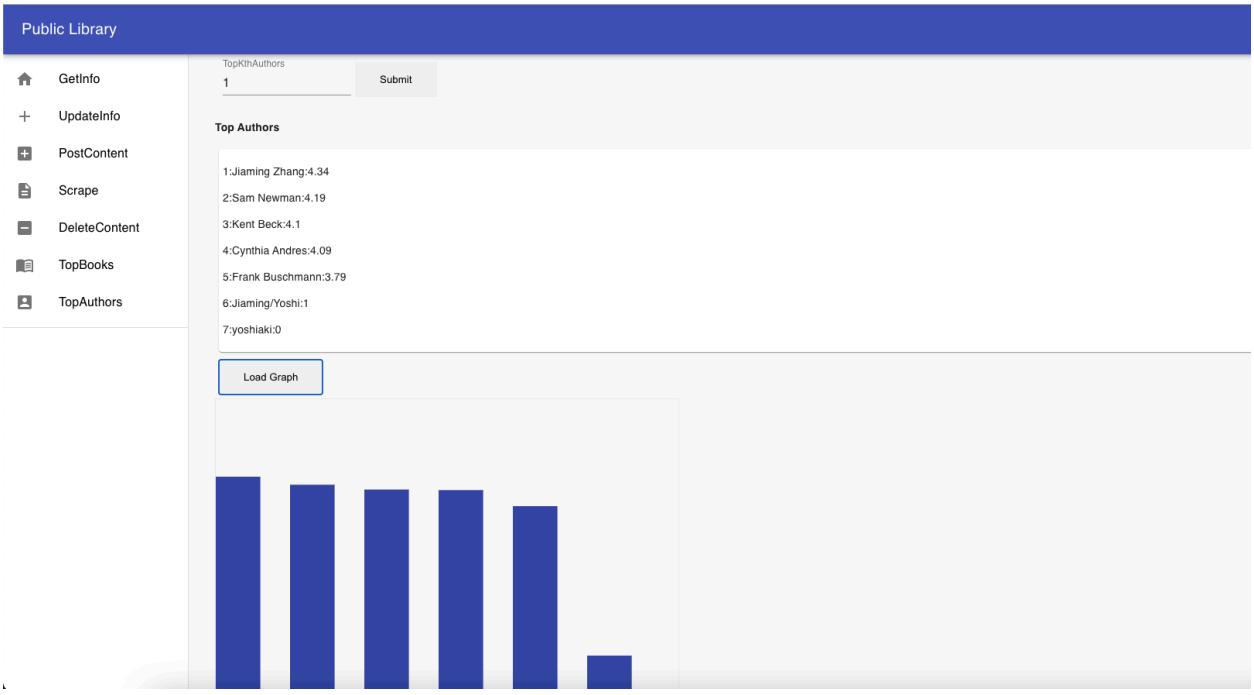


The overall appearance

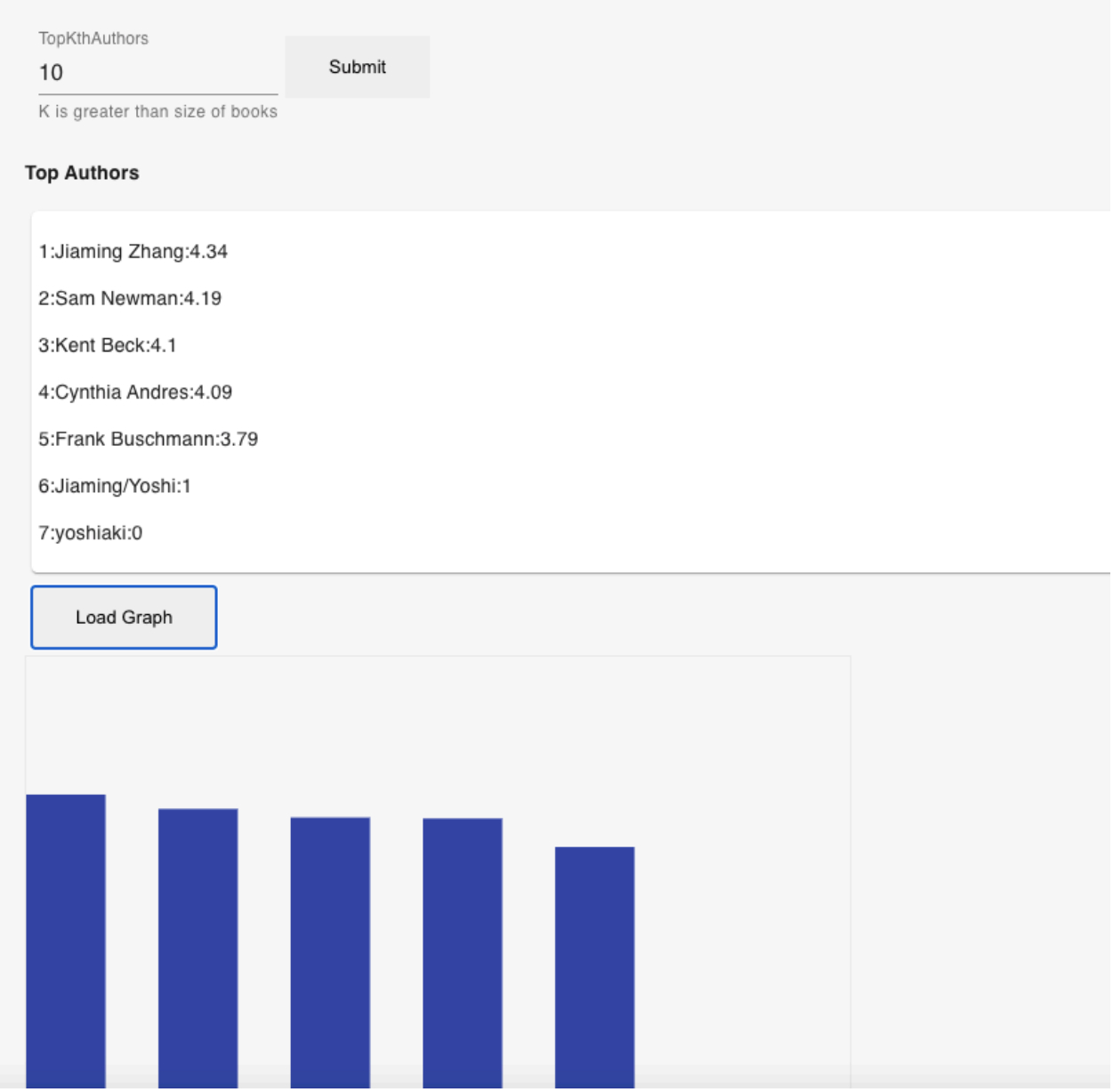
Graph when reset k = 2



10. Top Authors Graph



Normal



Exceeding max size, normally display all authors