

Manual Test Plan For Web Scrapping 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=HhMDV8vMa5&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"
```

```

try
    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:DesigningFin
e-GrainedSystems\n': '22512931', '\nAccelerate:BuildingandScalingHigh-PerformingTechnologyOrganizati
ons\n': '35747076', '\nReleaseIt!:DesignandDeployProduction-ReadySoftware(PragmaticProgrammers)\n':
'1069827', '\nFundamentalsofSoftwareArchitecture:AnEngineeringApproach\n': '44144493', '\nThePragmat
icProgrammer:FromJourneymentoMaster\n': '4099', 'IamCSStudent': '67833', 'Jiaming Zhangs Book': '123
4', '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-ReadySoftw
are(PragmaticProgrammers)\n': '1069827', '\nFundamentalsofSoftwareArchitecture:AnEngineeringApproach
\n': '44144493', '\nThePragmaticProgrammer:FromJourneymentoMaster\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/author/show/38241.Cynthia_Andres', 'Sam Newman': 'https://www.goodreads.com/author/show/3362774.Sam_Newman', 'Nicole Forsgren': 'https://www.goodreads.com/author/show/17037914.Nicole_Forsgren', 'Jez Humble': 'https://www.goodreads.com/author/show/4149510.Jez_Humble', 'Gene Kim': 'https://www.goodreads.com/author/show/328437.Gene_Kim', 'Michael T. Nygard': 'https://www.goodreads.com/author/show/6089.Michael_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.goodreads.com/author/show/2815.Andy_Hunt', 'Dave Thomas': 'https://www.goodreads.com/author/show/13.Dave_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 21"
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_author
    res = complex_author_query(query)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 66, in complex_author_query
    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_author
    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.

```
(base) JoeldeMacBook-Pro:sp21-cs242-assignment2 joelzhang$ python main.py -aq "author.author_id:AND
2"
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_auth
thor
    res = complex_author_query(query)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 61, in complex_author_q
query
    res = not_query_author(op, value, attr)
  File "/Users/joelzhang/cs242/sp21-cs242-assignment2/query_author.py", line 109, in not_query_autho
r
    raise ValueError("Invalid input")
ValueError: Invalid input
```

Visualization Single Web App Manual Test Plan

1. General Main Page

The page appearance should be as first opened:

The screenshot displays the 'Public Library' web application interface. It features a dark blue header bar with the text 'Public Library'. On the left side, there is a vertical sidebar containing a list of navigation links, each preceded by a small icon: 'GetInfo' (house icon), 'UpdateInfo' (plus icon), 'PostContent' (plus icon in a square), 'Scrape' (document icon), 'DeleteContent' (minus icon), 'TopBooks' (book icon), and 'TopAuthors' (person icon). The main content area on the right is light gray and contains two identical search forms stacked vertically. Each form has a 'Select' dropdown menu with 'id' selected, a text input field containing 'BookId' or 'AuthorId', and a 'Submit' button. Below the first form, there is a message: 'Please select search key'. Below the second form, there is a message: 'Please select search key'. At the bottom of the main content area, there is a message: 'No data currently, Please reinput the field and check value validation'.

2. Responsive appearance on mobile device:

Public Library

- GetInfo
- UpdateInfo
- PostContent
- Scrape
- DeleteContent
- TopBooks
- TopAuthors

Select
id
Please select search key

BookId

Submit

Select
id
Please select search key

AuthorId

Submit

No data currently. Please reinput the field and check value validation

Appearance on iPhone 5 emulator.

3. Successfully get book by id

Select

id

▼

BookId

Please select search key

Submit

Select

id

▼

AuthorId

Please select search key

Submit

_id => 19990225
book_author => JiamingWang
book_id => 19990225
book_rating => 5
book_rating_count => 1203
book_review_count => 1255
book_title => NewValue
book_url => www.jmz.com
isbn => 121

4. Failed to get book from database

Select

id

▼

BookId

Please select search key

Submit

Select

id

▼

AuthorId

Please select search key

Submit

No data currently, Please reinput the field and check value validation

5. Successfully get author by id:

Select

id

▼

AuthorId

Submit

Please select search key

_id => 123124

author_book => jiamingyoshi

author_id => 123124

author_name => yoshiaki

author_rating => 0

author_rating_counts => 0

author_review_counts => 0

author_url => www.google.com

6. Update book url by id

BookId	Select	UpdateBook	Submit
19990225	URL	www.hoohoo.com	
Please select search key			

Please select search key

Success Put!

Success information output

BookId	Select	UpdateBook	Submit
	URL		
Please select search key			

AuthorId	Select	UpdateAuthor	Submit
Please select search key			

Success Put!

valid key update for book:
book_rating, isbn, book_title, book_url, book_review_count, book_rating_count.
valid key update for author:
author_name, author_rating, author_rating_counts, author_review_counts, author_url

Overall page

Failed to put, please check format and key of inputing value Or the field to update is not existing in the data structure

valid key update for book:
book_rating, isbn, book_title, book_url, book_review_count, book_rating_count.
valid key update for author:
author_name, author_rating, author_rating_counts, author_review_counts, author_url

Warning and hint information

7. Post

Overall Appearance:

BookId	BookAuthor	BookRating 0	BookRatingCount 0	BookReviewCount 0	BookTitle	BookURL	Submit
AuthorId	AuthorBook	AuthorRating 0	AuthorRatingCount 0	AuthorReviewCount 0	AuthorName	AuthorURL	Submit

Nothing to be posted

In multi book authors situation, please follow format of author1/author2/... to insert authors
In multi author books situation, please follow same format as book authors, separate with "/"
All rating and counts should be positive, 0 otherwise input.

Success Post!

Success inform

8. Scrape and insert to database

ScrapeURL

Submit

Failed to post, please check format and key of inputing value

The correct format example usage:
book/show/xxx.BookName
author/show/xxx.AuthorName
You dont need goodreads prefix

Failed to get and hint information

ScrapeURL

Submit

Success Post!

Success Posting

9. Delete a book by id

Failed

BookId

Submit

AuthorId

Submit

Deleted failed, check your if id is valid

BookId

Submit

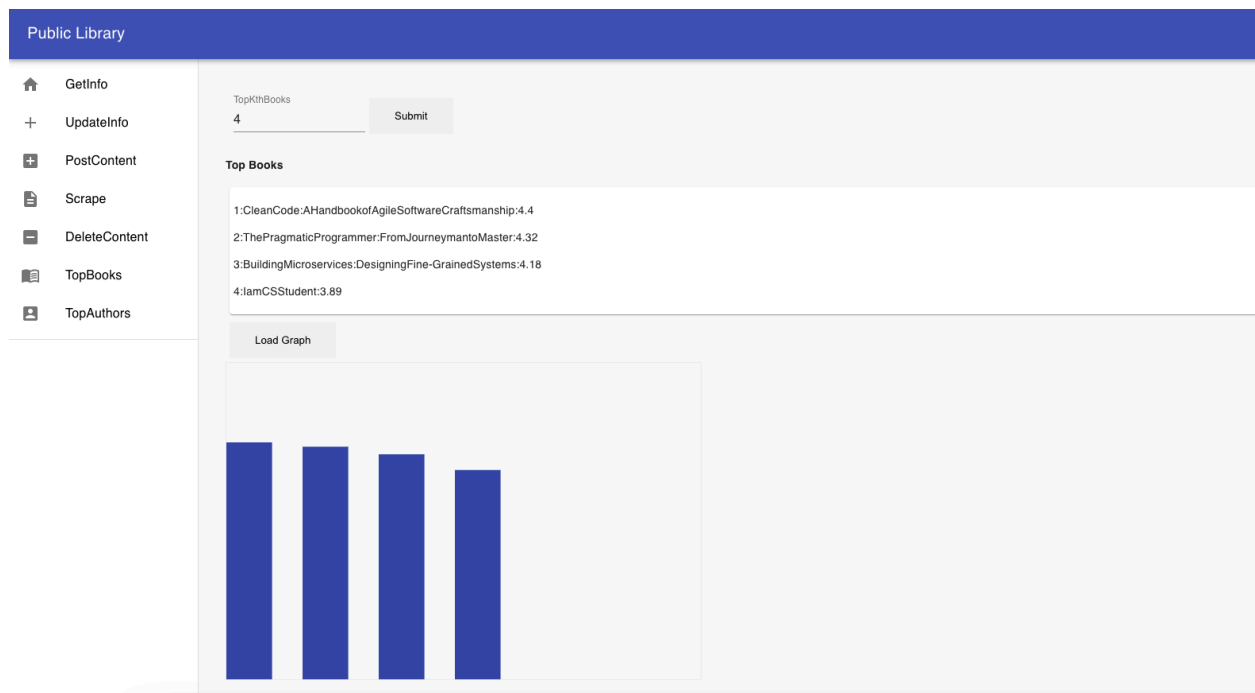
AuthorId

Submit

Success Delete!

Success

9. Top Books Visual



TopKthBooks

4

Submit

Top Books

1:CleanCode:AHandbookofAgileSoftwareCraftsmanship:4.4

2:ThePragmaticProgrammer:FromJourneytoMaster:4.32

3:BuildingMicroservices:DesigningFine-GrainedSystems:4.18

4:IamCSSStudent:3.89

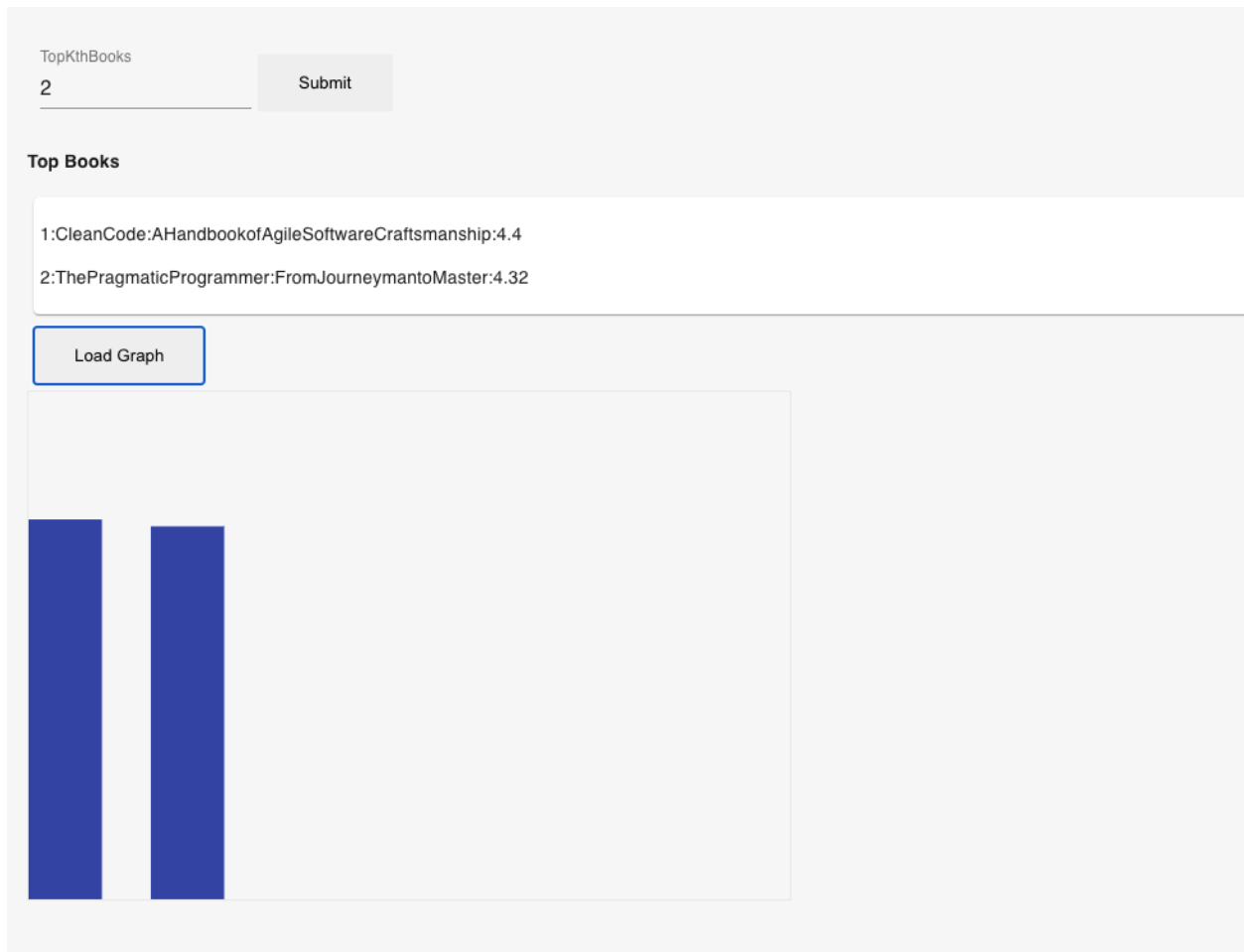
Load Graph



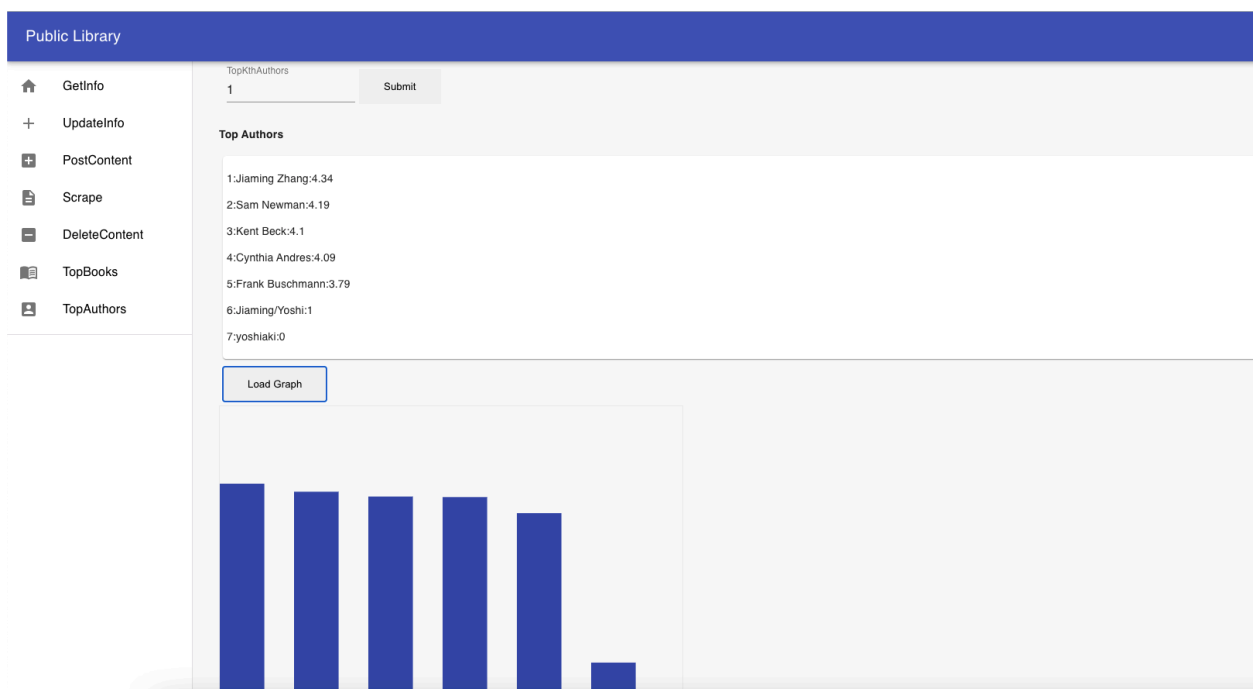
Rank	Book Title	Score
1	CleanCode:AHandbookofAgileSoftwareCraftsmanship	4.4
2	ThePragmaticProgrammer:FromJourneytoMaster	4.32
3	BuildingMicroservices:DesigningFine-GrainedSystems	4.18
4	IamCSSStudent	3.89

The overall appearance

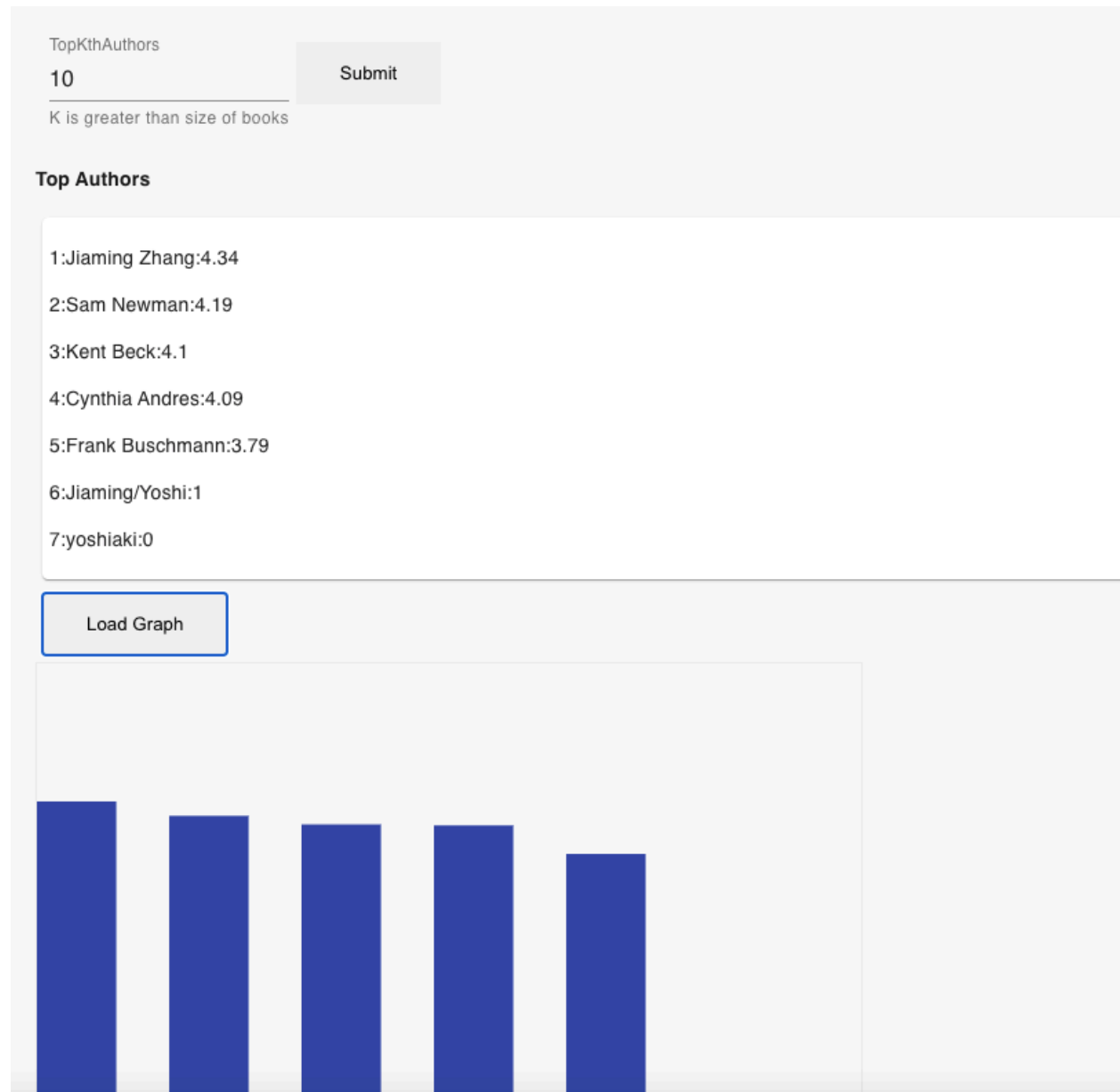
Graph when reset $k = 2$



10. Top Authors Graph



Normal



Exceeding max size, normally display all authors