NYIT

College of Engineering & Computing Sciences

Book Recommendation System

Timeline: October 5th - December 15th

Xiaohui Chen - Rajendra Bhagroo - Shelly Gupta

Problem Statement

Problem Statement

- Online recommendation systems have become a trend in recent history.

- Recommendation systems are broadly used to recommend specific products to end users based on user statistics.

- A recommendation system is one of the strongest tools to increase profits and retain customers and increase customer satisfaction.

- The existing systems lead to extraction of irrelevant information and lead to lack of user satisfaction.

Model

Data preprocessing

BX-Books.csv, BX-Users.csv, BX-Book-Ratings.csv

Dataset we use: Collected by Cai-Nicolas Ziegler in 2004 from August - September from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. Contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books.

http://www2.informatik.uni-freiburg.de/~cziegler/BX/

- Drop image urls of book dataset
- Replace nAn for string-type values with ‘unknown’ after double check
- Check data validity & fix data type problems
  - i.e. yearOfPublication contains some non-integer values, values for different columns were misplaced.
- Fix logical data problems
  - i.e. yearOfPublication should be less than 2004 (dataset publication year) and greater than 0
  - i.e. age of users should be within a reasonable range (5, 90)
- Replace nAn for integer-type values with the arithmetic mean of other values in that column
- Fix data consistency problems
  - i.e. all books exist in the rating dataset should also exist in the book dataset
  - I.e. all users exist in the rating dataset should also exist in the user dataset

Algorithm

Collaborative Filtering using Euclidean Distance

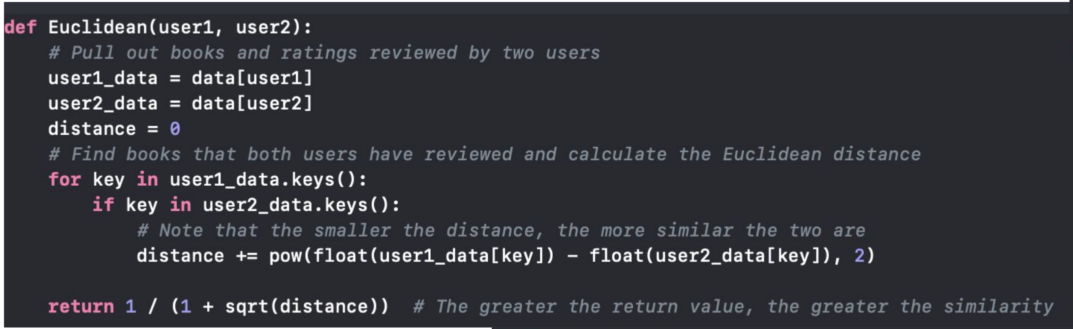
- Core idea of collaborative filtering
- Compute Distance-based Similarity
    - Euclidean Distance: the square root of the sum of squared differences between corresponding elements of the two vectors.

Inversely Proportional

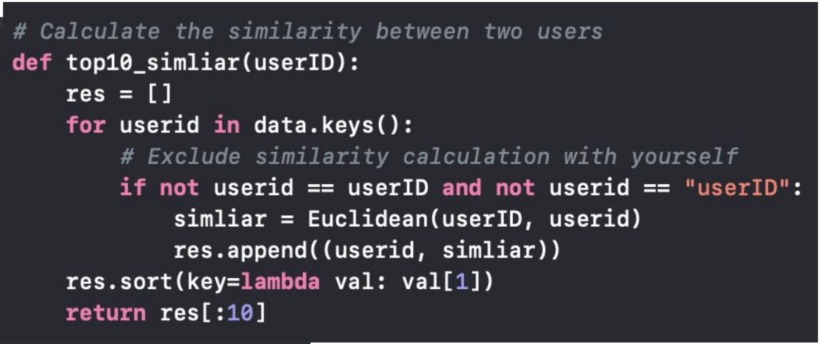
- Similarity: $$\frac{1}{1 + d(p_1, p_2)}$$

Code

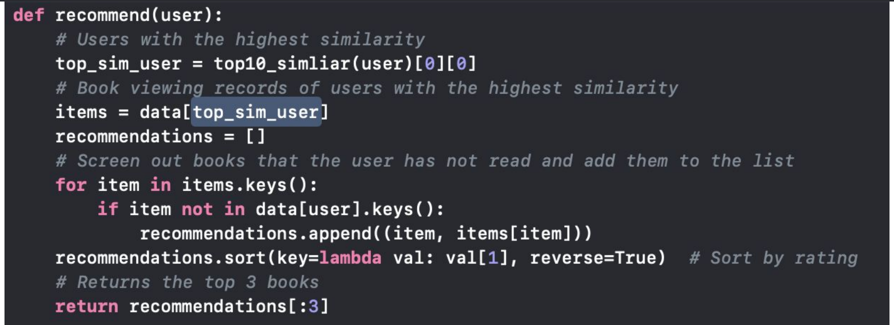
books = pd.read_csv('BX-Books.csv', sep=';', error_bad_lines=False, encoding="cp1252")
users = pd.read_csv('BX-Users.csv', sep=';', error_bad_lines=False, encoding="cp1252")
ratings = pd.read_csv('BX-Book-ratings.csv', sep=';', error_bad_lines=False, encoding="cp1252")
--- After Data Cleaning --
data = pd.merge(books, ratings, on='ISBN')
data[['userID', 'bookRating', 'ISBN', 'bookTitle']].sort_values('userID').to_csv('data.csv', index=False)
* data contains the book and rating for each user
--- Computation---



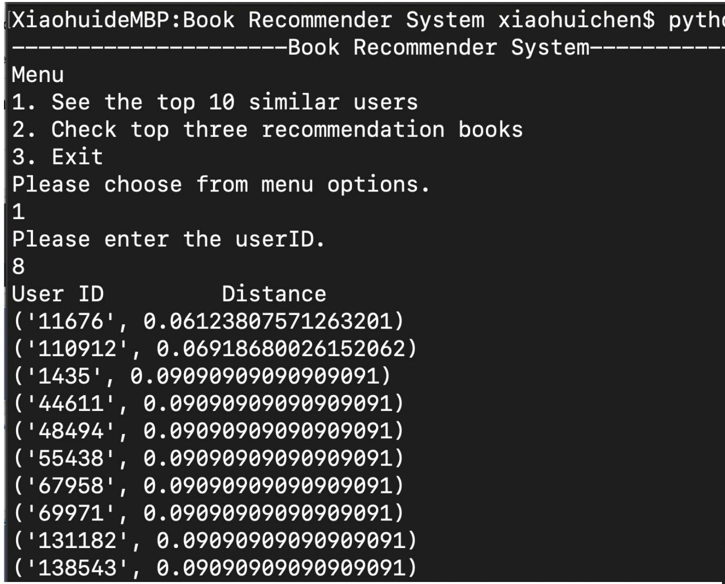
```
def Euclidean(user1, user2):
    # Pull out books and ratings reviewed by two users
    user1_data = data[user1]
    user2_data = data[user2]
    distance = 0
    # Find books that both users have reviewed and calculate the Euclidean distance
    for key in user1_data.keys():
        if key in user2_data.keys():
            # Note that the smaller the distance, the more similar the two are
            distance += pow(float(user1_data[key]) - float(user2_data[key]), 2)

    return 1 / (1 + sqrt(distance))  # The greater the return value, the greater the similarity
```



```
# Calculate the similarity between two users
def top10_simliar(userID):
    res = []
    for userid in data.keys():
        # Exclude similarity calculation with yourself
        if not userid == userID and not userid == "userID":
            simliar = Euclidean(userID, userid)
            res.append((userid, simliar))
    res.sort(key=lambda val: val[1])
    return res[:10]
```



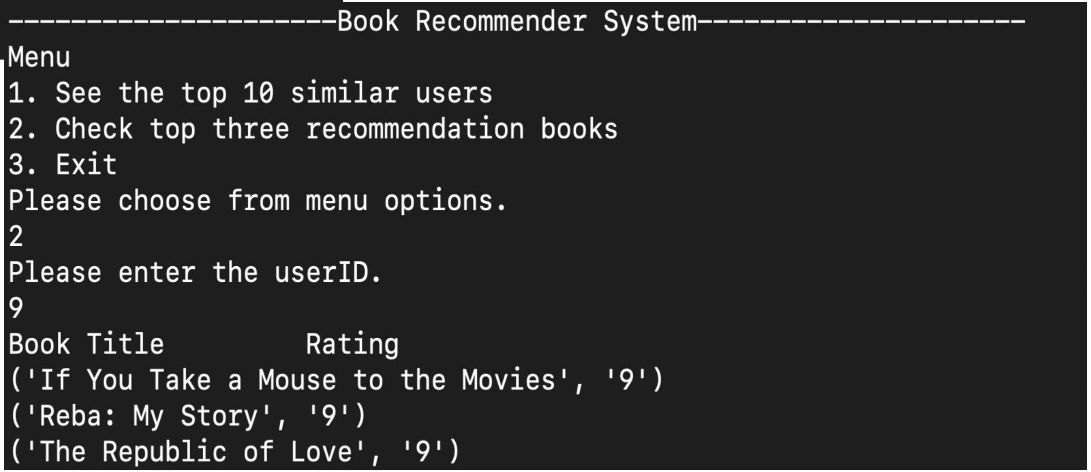
```
def recommend(user):
    # Users with the highest similarity
    top_sim_user = top10_simliar(user)[0][0]
    # Book viewing records of users with the highest similarity
    items = data[top_sim_user]
    recommendations = []
    # Screen out books that the user has not read and add them to the list
    for item in items.keys():
        if item not in data[user].keys():
            recommendations.append((item, items[item]))
    recommendations.sort(key=lambda val: val[1], reverse=True)  # Sort by rating
    # Returns the top 3 books
    return recommendations[:3]
```



```
XiaohuideMBP:Book Recommender System xiaohuichen$ python bookRecommender.py
---------------------Book Recommender System---------------------
Menu
1. See the top 10 similar users
2. Check top three recommendation books
3. Exit
Please choose from menu options.
1
Please enter the userID.
8
User ID         Distance
('11676', 0.061238075712632O1)
('110912', 0.06918680026152062)
('1435', 0.09090909090909091)
('44611', 0.09090909090909091)
('48494', 0.09090909090909091)
('55438', 0.09090909090909091)
('67958', 0.09090909090909091)
('69971', 0.09090909090909091)
('131182', 0.09090909090909091)
('138543', 0.09090909090909091)
```

See top10 similar users

See top3 recommendation books

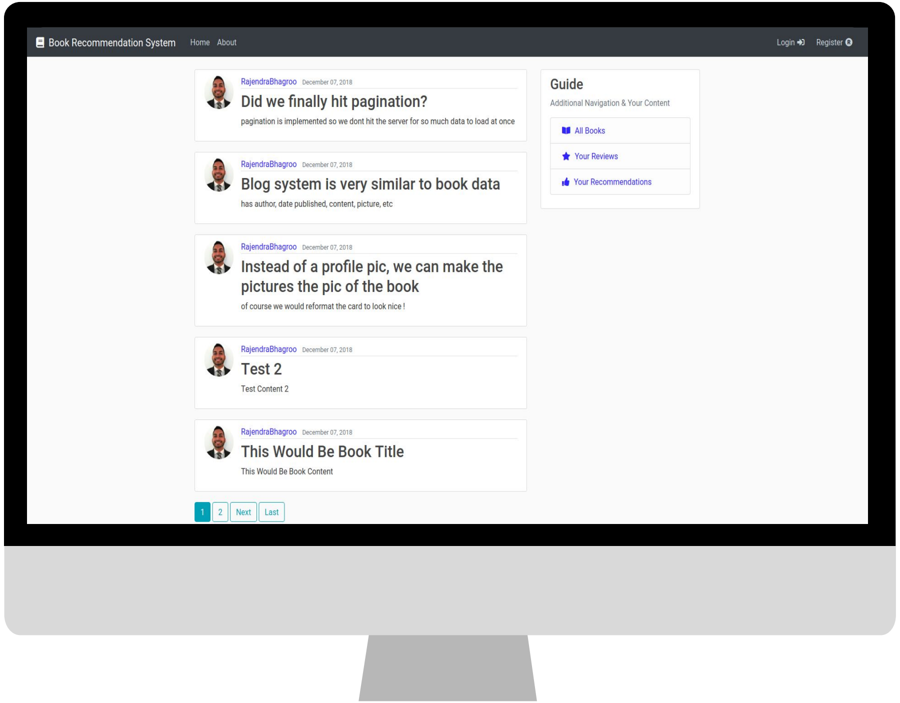


```
---------------------Book Recommender System---------------------
Menu
1. See the top 10 similar users
2. Check top three recommendation books
3. Exit
Please choose from menu options.
2
Please enter the userID.
9
Book Title         Rating
('If You Take a Mouse to the Movies', '9')
('Reba: My Story', '9')
('The Republic of Love', '9')
```

User Interface

User Interface

Django



Book Recommendation System Home About Login Register

RajendraBhagroo December 07, 2018

Did we finally hit pagination?

pagination is implemented so we dont hit the server for so much data to load at once

RajendraBhagroo December 07, 2018

Blog system is very similar to book data

has author, date published, content, picture, etc

RajendraBhagroo December 07, 2018

Instead of a profile pic, we can make the pictures the pic of the book

of course we would reformat the card to look nice !

RajendraBhagroo December 07, 2018

Test 2

Test Content 2

RajendraBhagroo December 07, 2018

This Would Be Book Title

This Would Be Book Content

1 2 Next Last

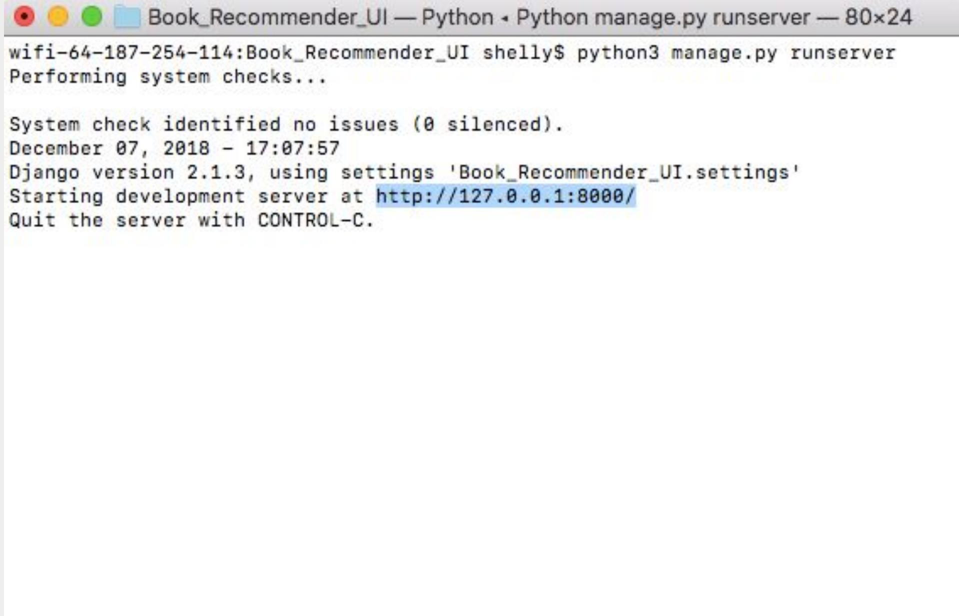
Guide

Additional Navigation & Your Content

- All Books
- Your Reviews
- Your Recommendations

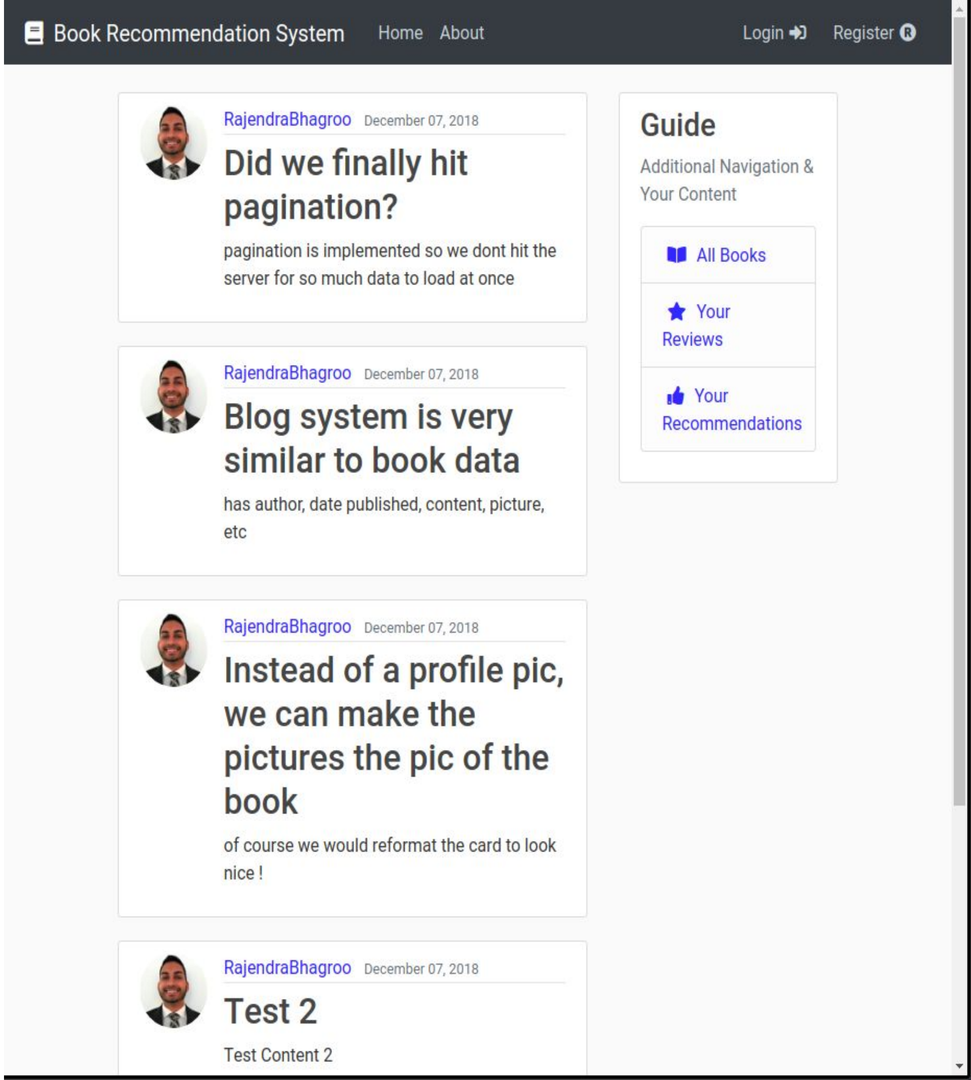
Our User Interface was developed in Python using the Django Framework

Running the Server


Book_Recommender_UI — Python ◂ Python manage.py runserver — 80×24

```
wifi-64-187-254-114:Book_Recommender_UI shelly$ python3 manage.py runserver
Performing system checks...

System check identified no issues (0 silenced).
December 07, 2018 - 17:07:57
Django version 2.1.3, using settings 'Book_Recommender_UI.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
```

Homepage Blog



RajendraBhagroo December 07, 2018

Did we finally hit pagination?

pagination is implemented so we dont hit the server for so much data to load at once

RajendraBhagroo December 07, 2018

Blog system is very similar to book data

has author, date published, content, picture, etc

RajendraBhagroo December 07, 2018

Instead of a profile pic, we can make the pictures the pic of the book

of course we would reformat the card to look nice !

RajendraBhagroo December 07, 2018

Test 2

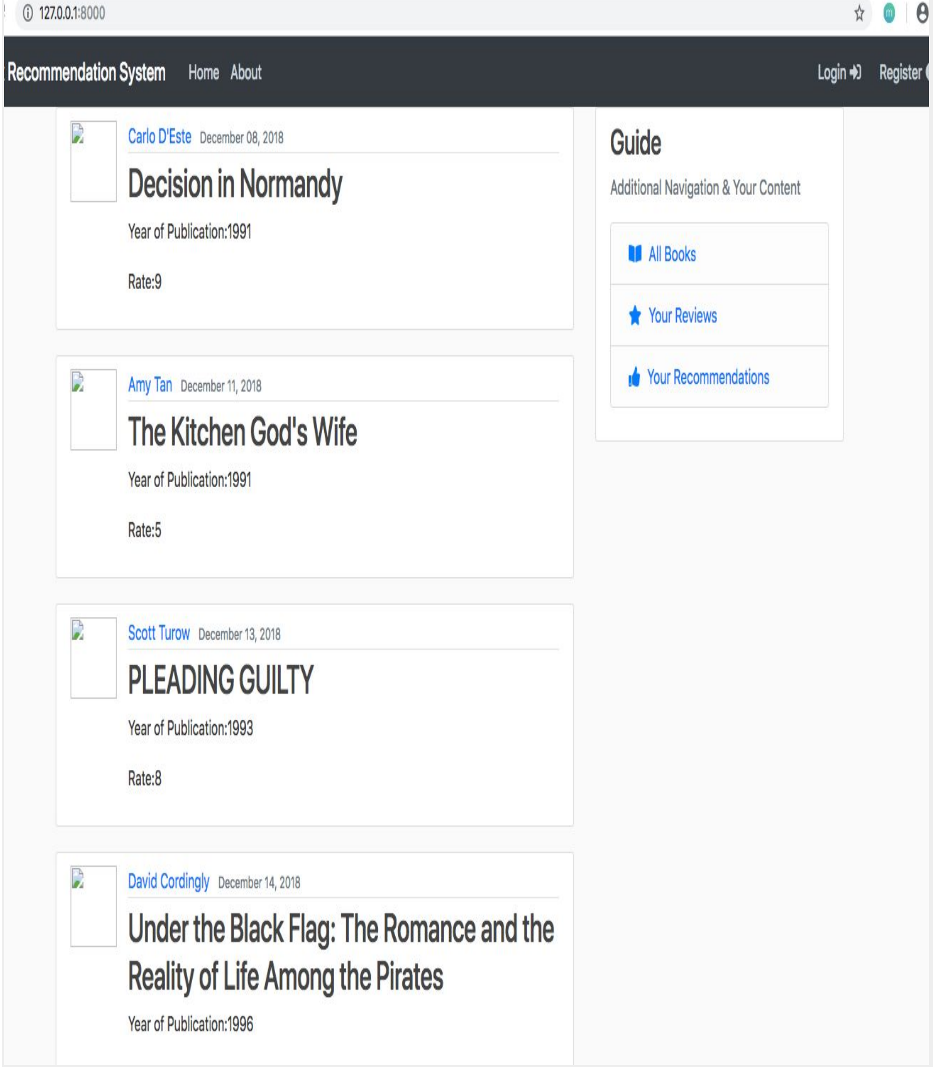
Test Content 2

Guide

Additional Navigation & Your Content

- All Books
- Your Reviews
- Your Recommendations

Homepage Books



Carlo D'Este December 08, 2018

Decision in Normandy

Year of Publication:1991

Rate:9

Amy Tan December 11, 2018

The Kitchen God's Wife

Year of Publication:1991

Rate:5

Scott Turow December 13, 2018

PLEADING GUILTY

Year of Publication:1993

Rate:8

David Cordingly December 14, 2018

Under the Black Flag: The Romance and the Reality of Life Among the Pirates

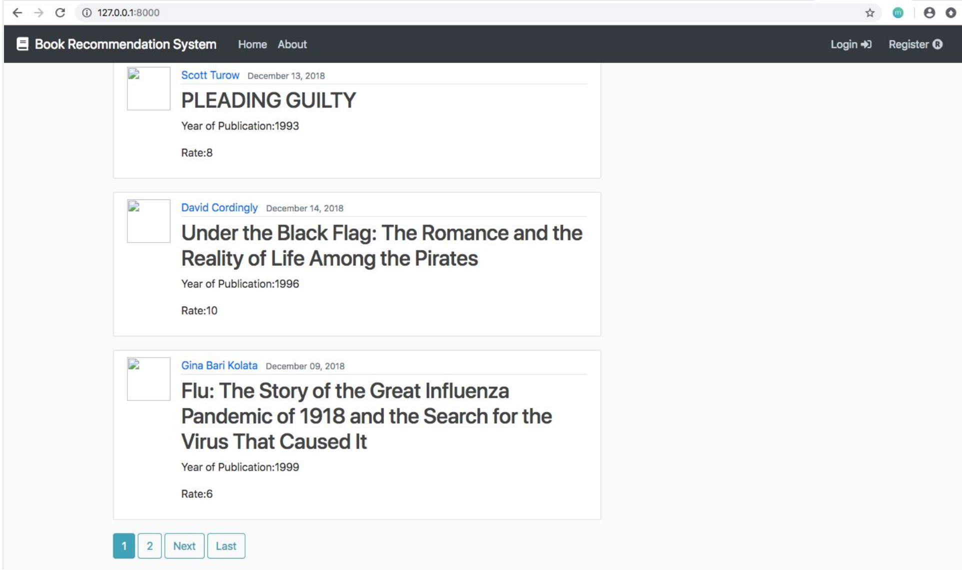
Year of Publication:1996

Guide

Additional Navigation & Your Content

- All Books
- Your Reviews
- Your Recommendations

Pagination



Scott Turow December 13, 2018

PLEADING GUILTY

Year of Publication:1993

Rate:8

David Cordingly December 14, 2018

Under the Black Flag: The Romance and the Reality of Life Among the Pirates

Year of Publication:1996

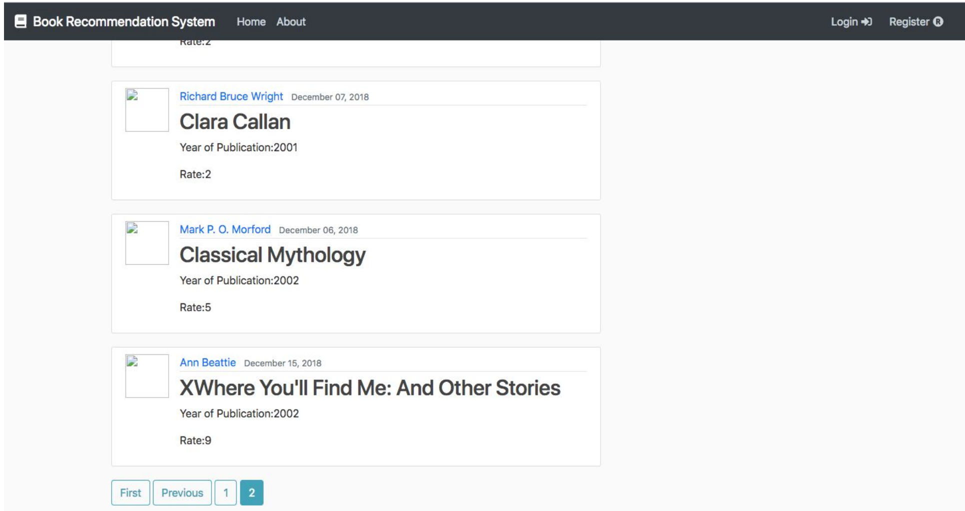
Rate:10

Gina Bari Kolata December 09, 2018

Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It

Year of Publication:1999

Rate:6



Rate:2

Richard Bruce Wright December 07, 2018

Clara Callan

Year of Publication:2001

Rate:2

Mark P. O. Morford December 06, 2018

Classical Mythology

Year of Publication:2002

Rate:5

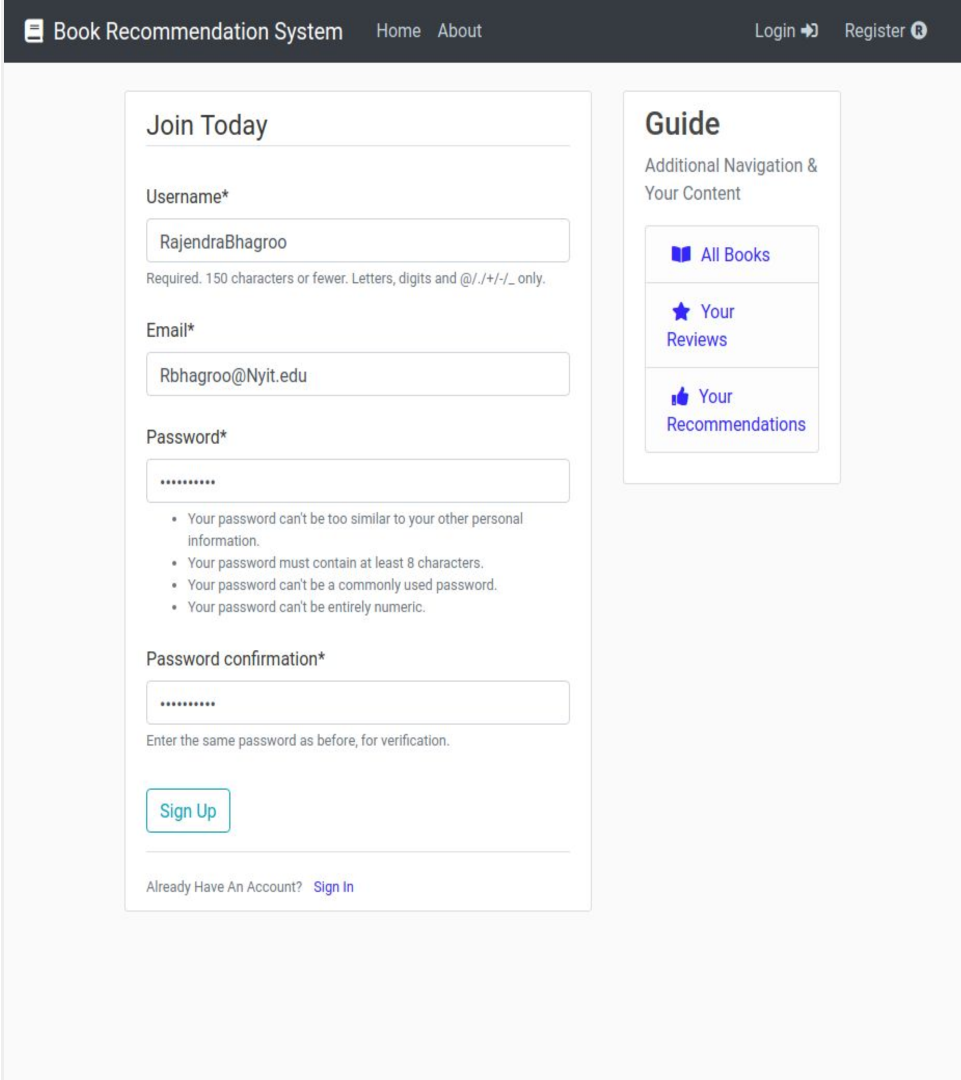
Ann Beattie December 15, 2018

XWhere You'll Find Me: And Other Stories

Year of Publication:2002

Rate:9

Register



Join Today

Username*

RajendraBhagroo

Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.

Email*

Rbhagroo@Nyit.edu

Password*

••••••••••

- Your password can't be too similar to your other personal information.
- Your password must contain at least 8 characters.
- Your password can't be a commonly used password.
- Your password can't be entirely numeric.

Password confirmation*

••••••••••

Enter the same password as before, for verification.

Sign Up

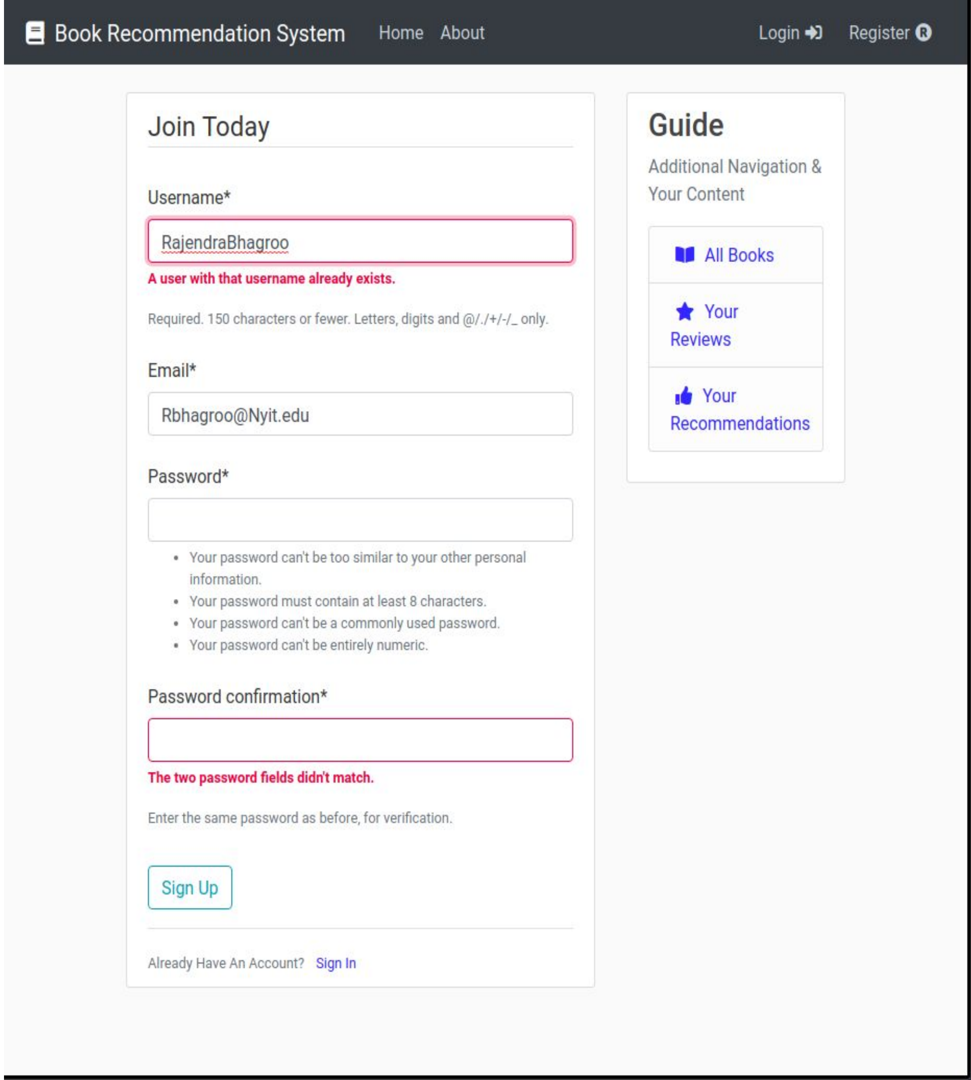
Already Have An Account? Sign In

Guide

Additional Navigation & Your Content

- All Books
- Your Reviews
- Your Recommendations

Register Verification



Join Today

Username*

RajendraBhagroo

A user with that username already exists.

Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.

Email*

Rbhagroo@Nyit.edu

Password*

- Your password can't be too similar to your other personal information.
- Your password must contain at least 8 characters.
- Your password can't be a commonly used password.
- Your password can't be entirely numeric.

Password confirmation*

The two password fields didn't match.

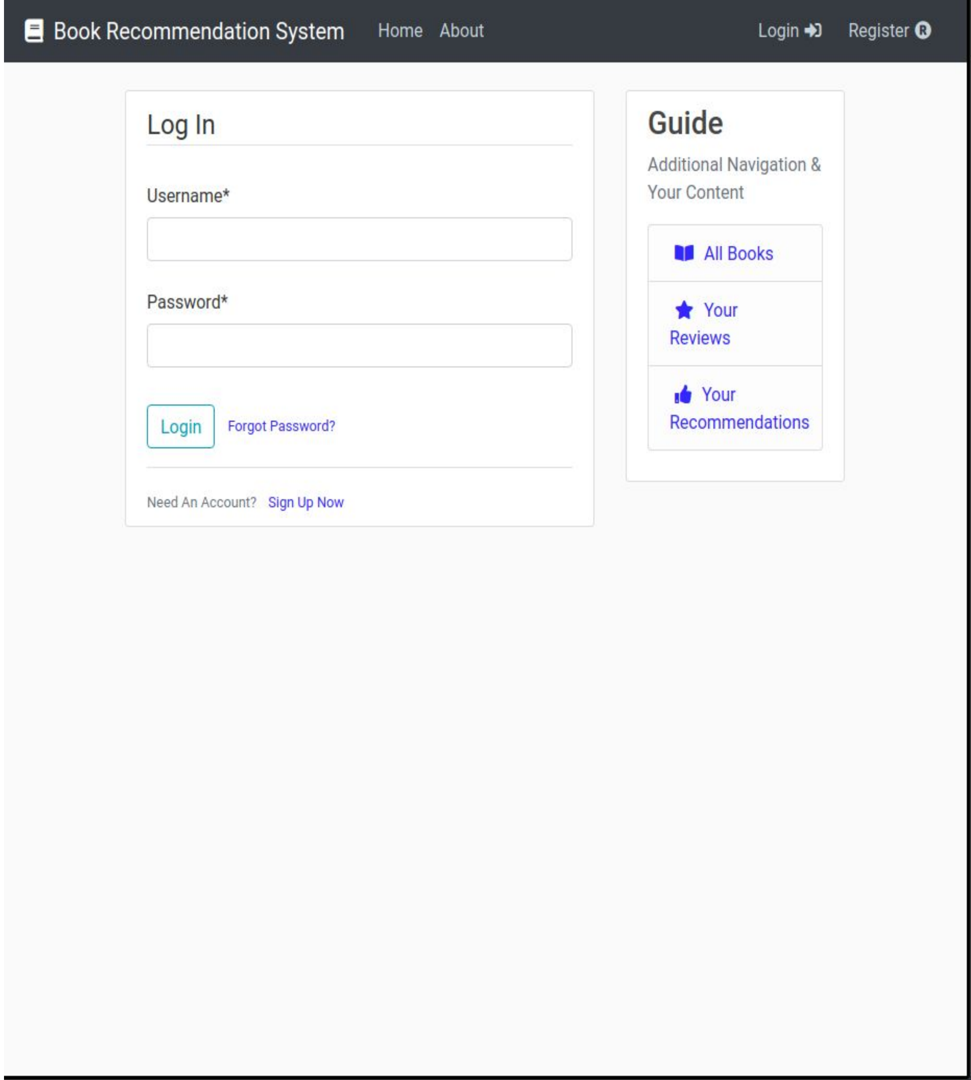
Enter the same password as before, for verification.

Already Have An Account? Sign In

Guide

Additional Navigation & Your Content

Login



Book Recommendation System

Log In

Username*

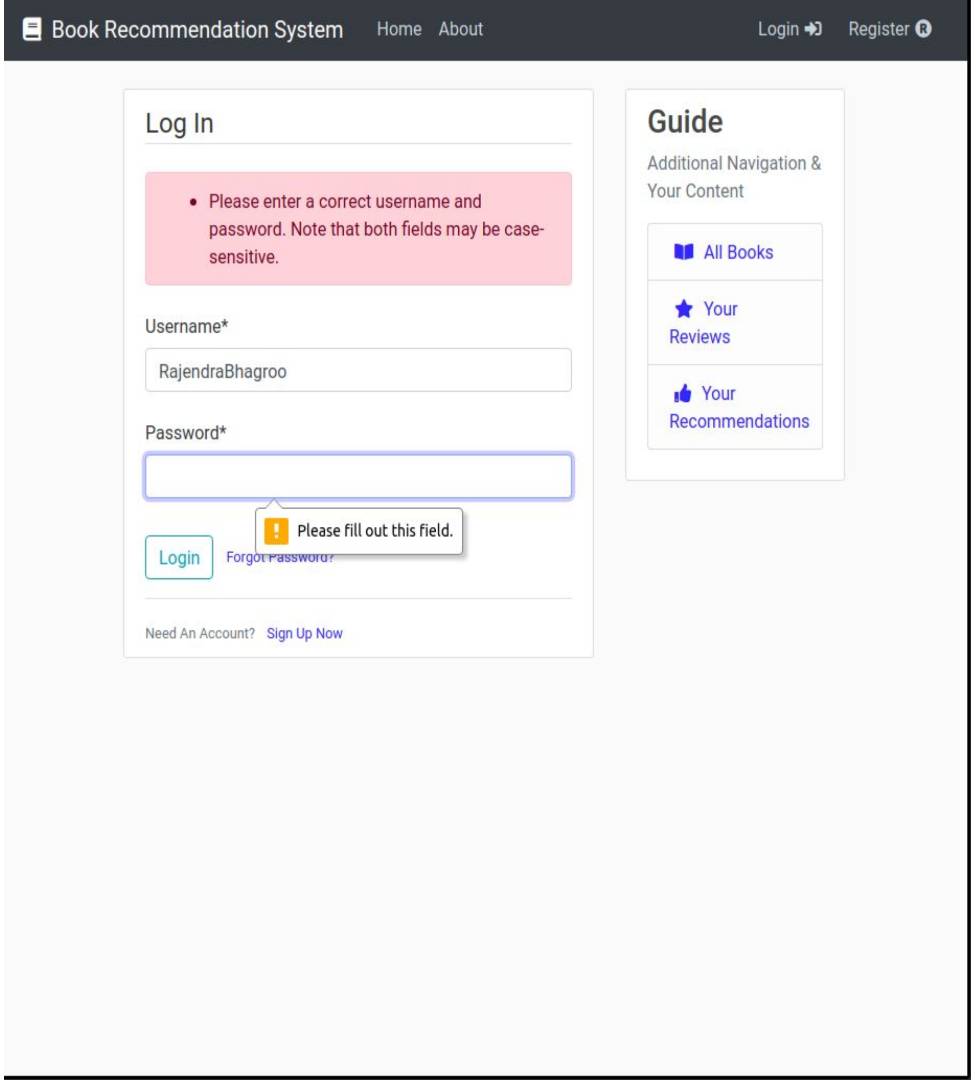
Password*

Need An Account? Sign Up Now

Guide

Additional Navigation & Your Content

Login Verification



Log In

- Please enter a correct username and password. Note that both fields may be case-sensitive.

Username*

RajendraBhagroo

Password*

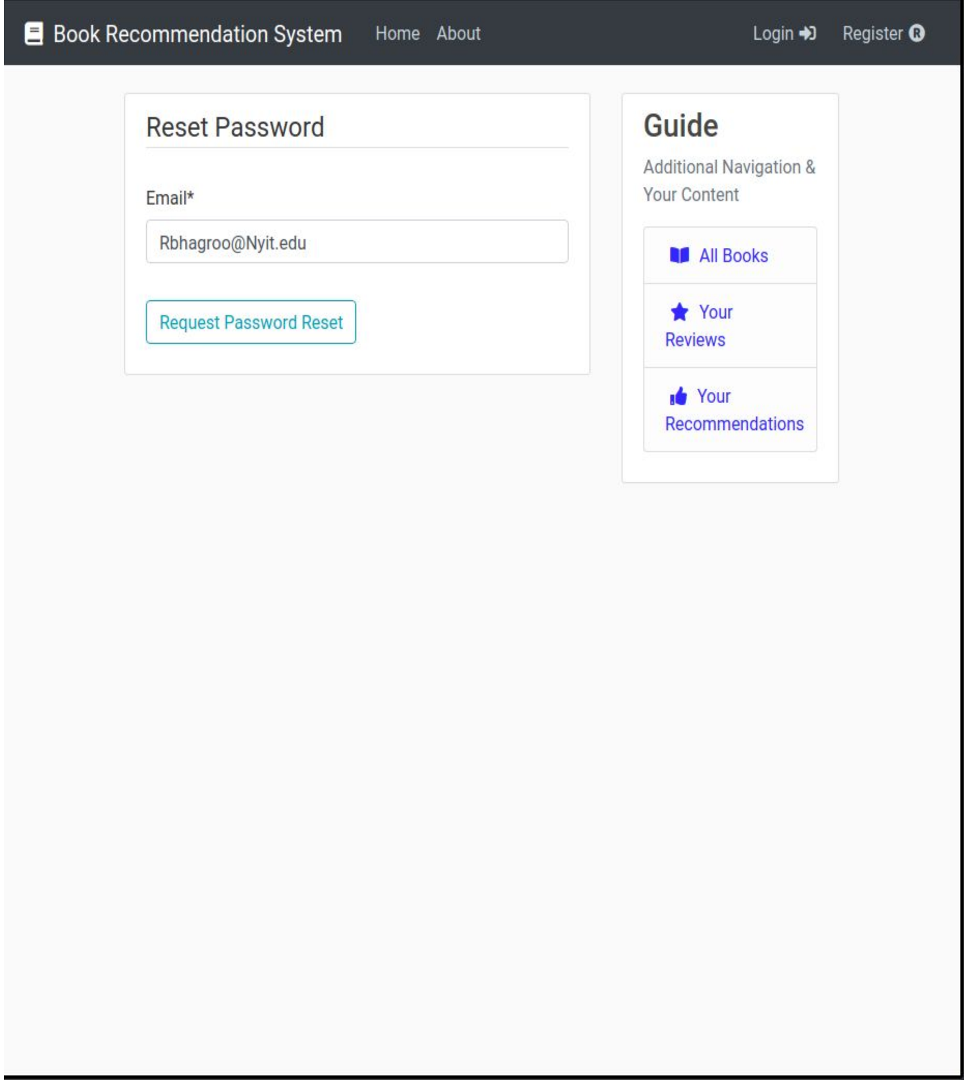
Please fill out this field.

Need An Account? Sign Up Now

Guide

Additional Navigation & Your Content

Password Reset



Reset Password

Email*

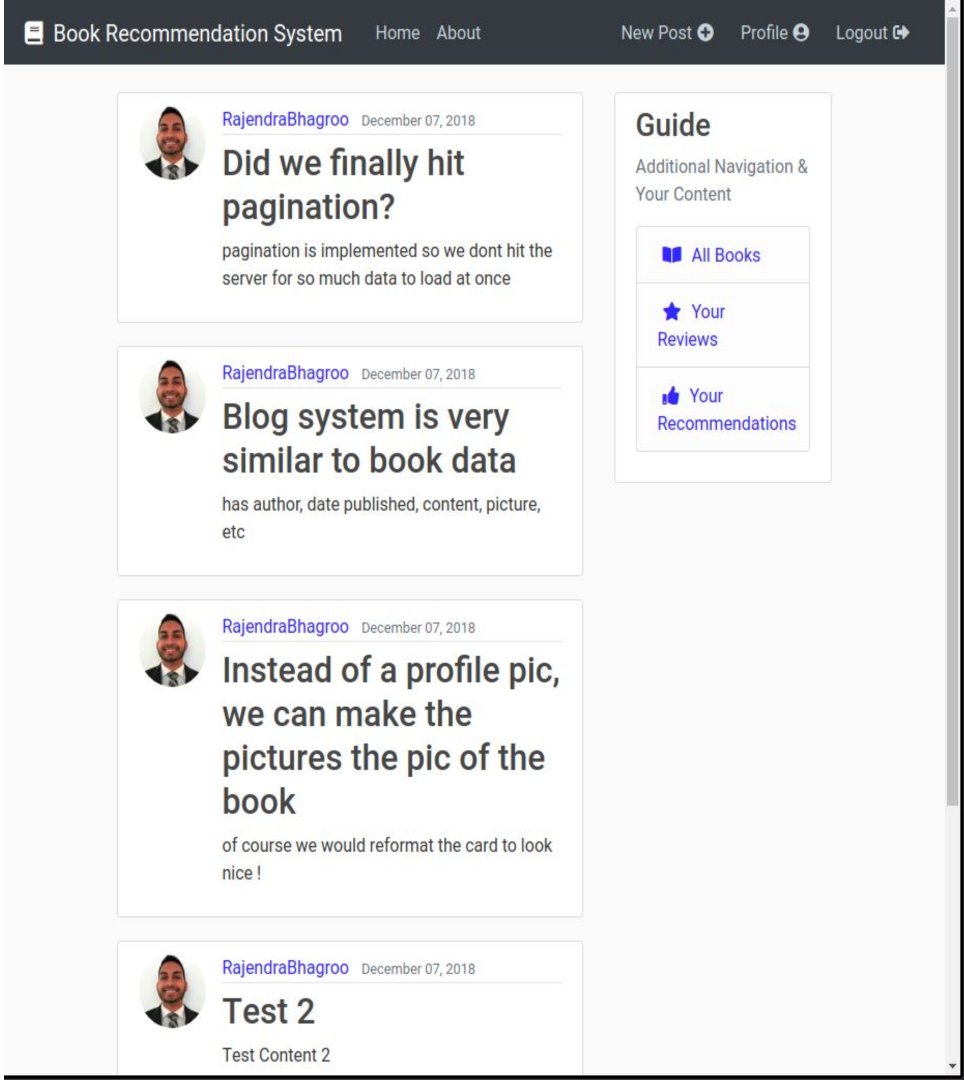
Rbhagroo@Nyit.edu

Guide

Additional Navigation & Your Content

Logged In

Note: Top Right Bar Changed



RajendraBhagroo December 07, 2018

Did we finally hit pagination?

pagination is implemented so we dont hit the server for so much data to load at once

RajendraBhagroo December 07, 2018

Blog system is very similar to book data

has author, date published, content, picture, etc

RajendraBhagroo December 07, 2018

Instead of a profile pic, we can make the pictures the pic of the book

of course we would reformat the card to look nice !

RajendraBhagroo December 07, 2018

Test 2

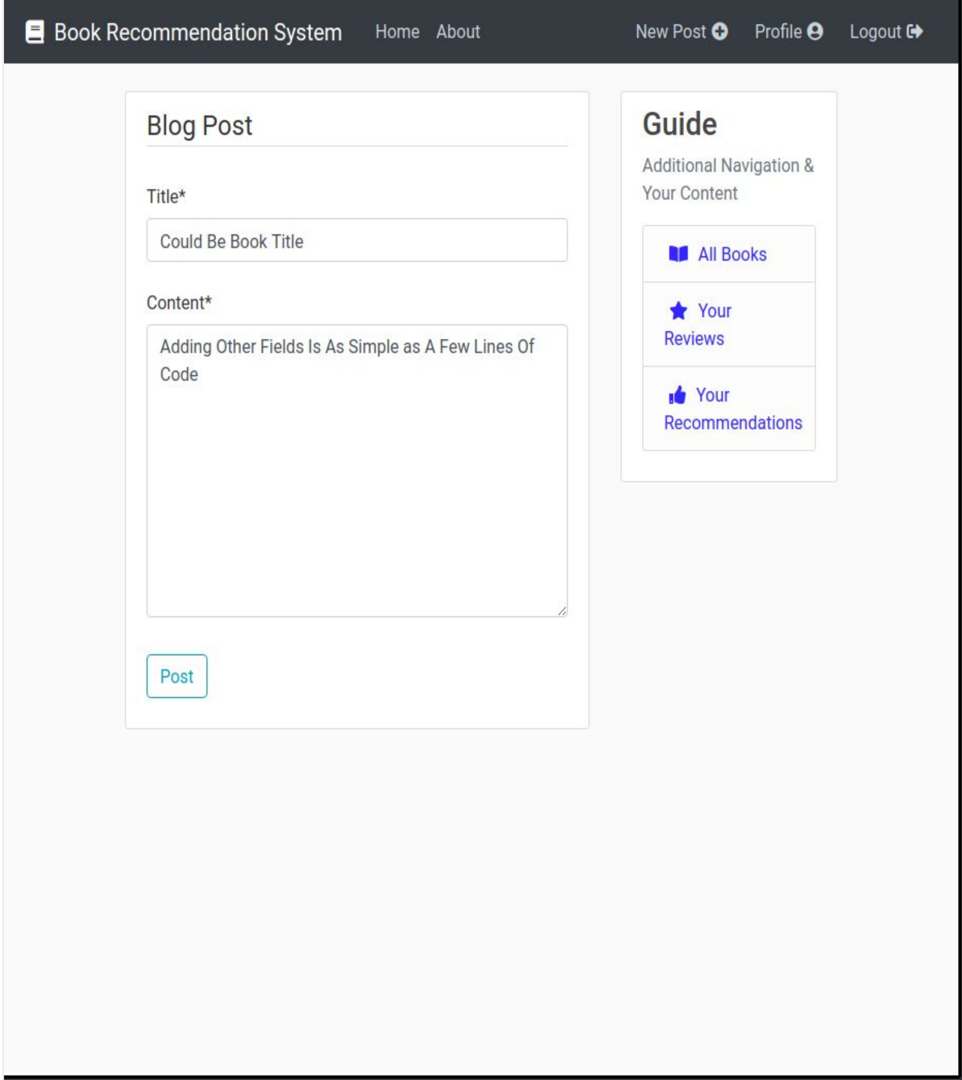
Test Content 2

Guide

Additional Navigation & Your Content

New Post

If this was book recommender, new book!



Blog Post

Title*

Could Be Book Title

Content*

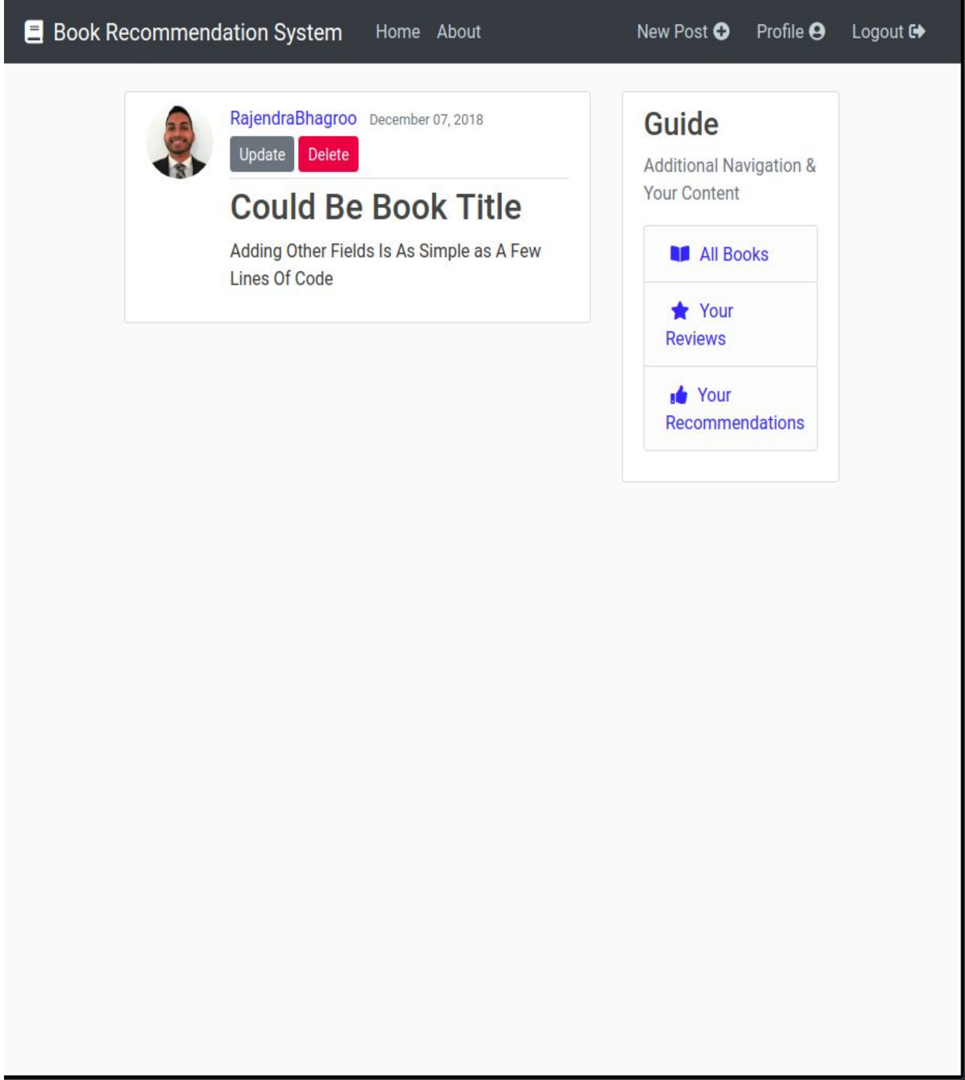
Adding Other Fields Is As Simple as A Few Lines Of Code

Guide

Additional Navigation & Your Content

Post
Update / Delete

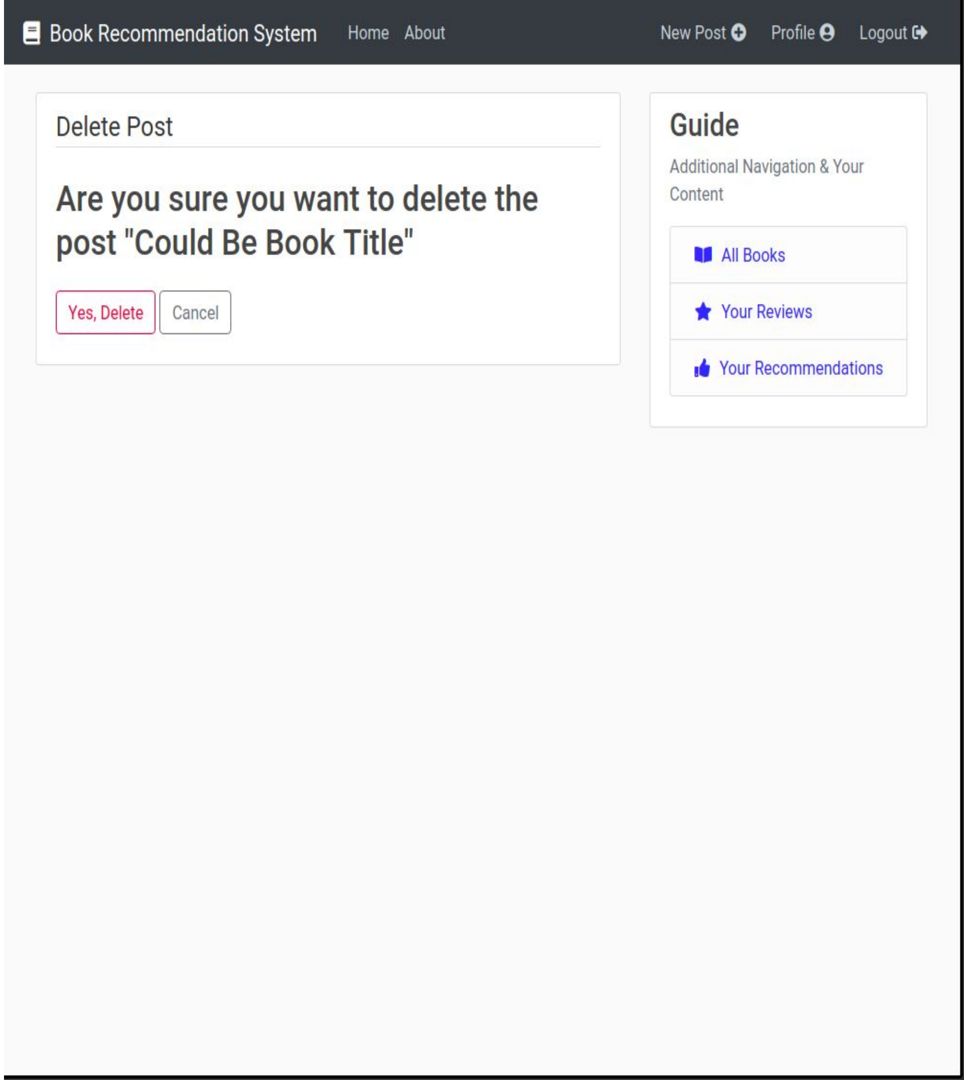
Only the original user can edit post when logged in



Could Be Book Title

Adding Other Fields Is As Simple as A Few Lines Of Code

Delete Post



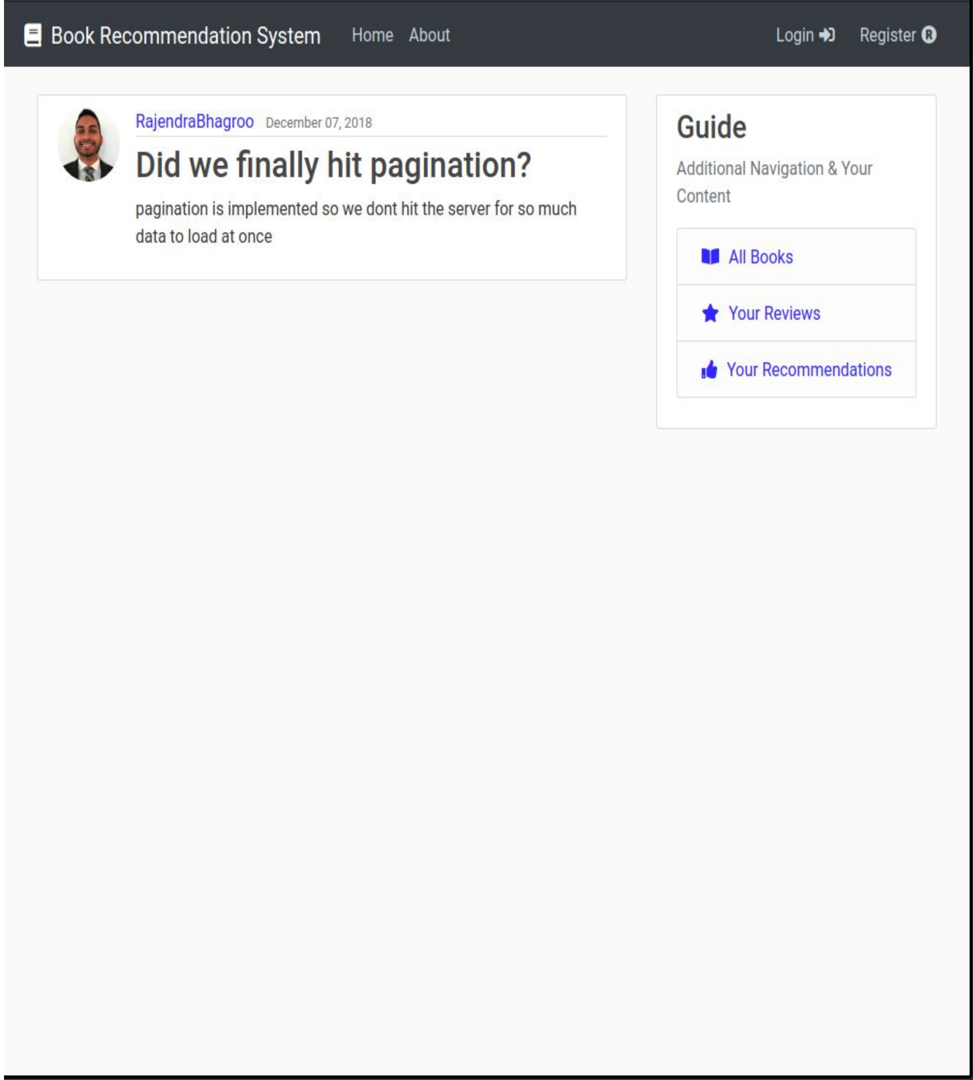
Delete Post

Are you sure you want to delete the post "Could Be Book Title"

Guide

Additional Navigation & Your Content

Logged Out Post

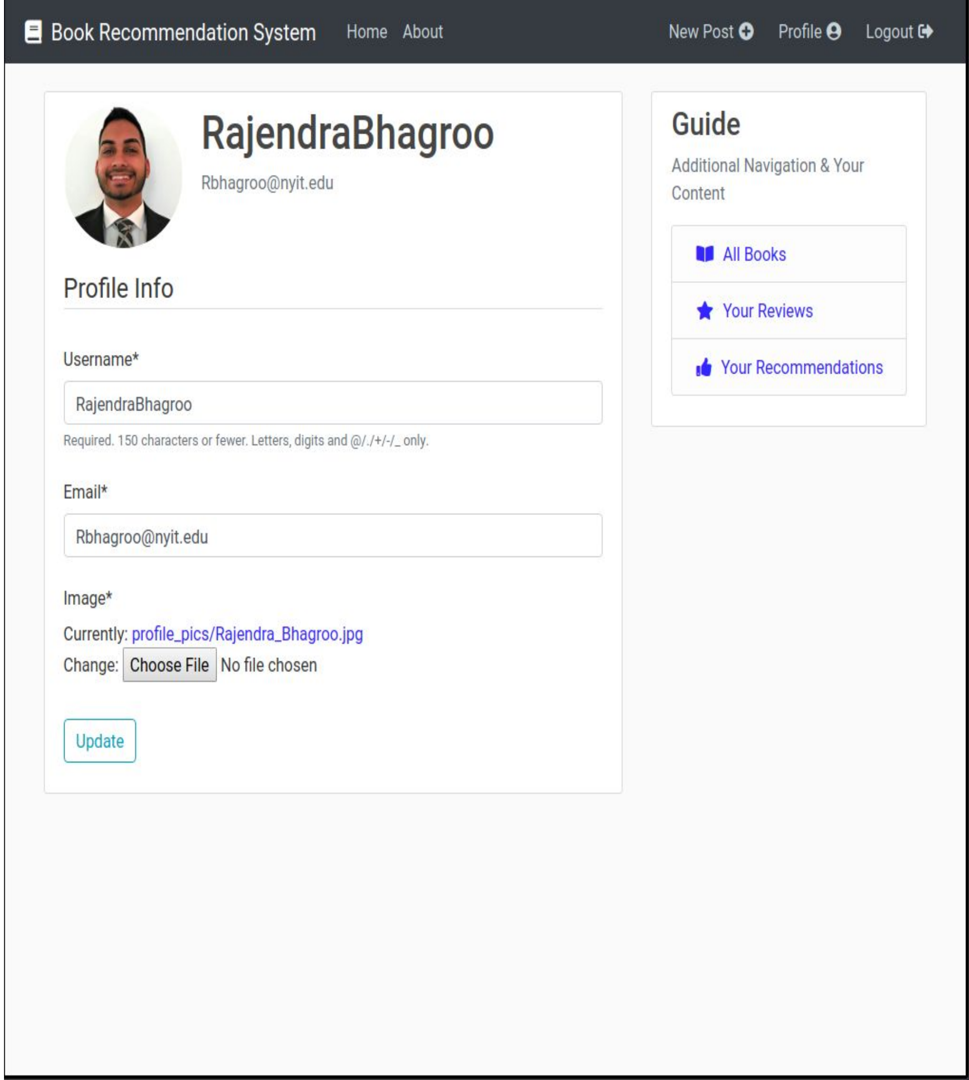


RajendraBhagroo December 07, 2018

Did we finally hit pagination?

pagination is implemented so we dont hit the server for so much data to load at once

Profile



RajendraBhagroo

Rbhagroo@nyit.edu

Profile Info

Username*

RajendraBhagroo

Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.

Email*

Rbhagroo@nyit.edu

Image*

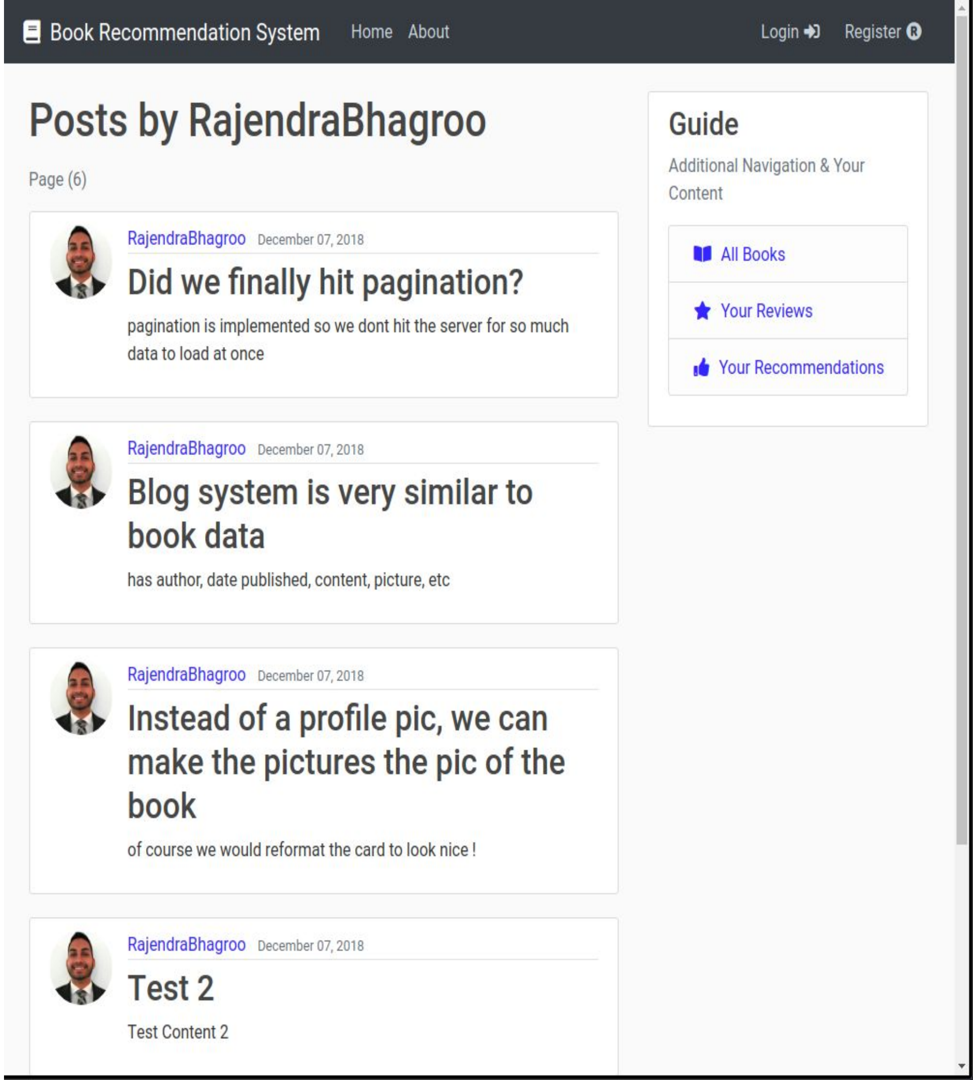
Currently: profile_pics/Rajendra_Bhagroo.jpg

Change: Choose File No file chosen

Guide

Additional Navigation & Your Content

Posts By User



Posts by RajendraBhagroo

Page (6)

RajendraBhagroo December 07, 2018

Did we finally hit pagination?

pagination is implemented so we dont hit the server for so much data to load at once

RajendraBhagroo December 07, 2018

Blog system is very similar to book data

has author, date published, content, picture, etc

RajendraBhagroo December 07, 2018

Instead of a profile pic, we can make the pictures the pic of the book

of course we would reformat the card to look nice !

RajendraBhagroo December 07, 2018

Test 2

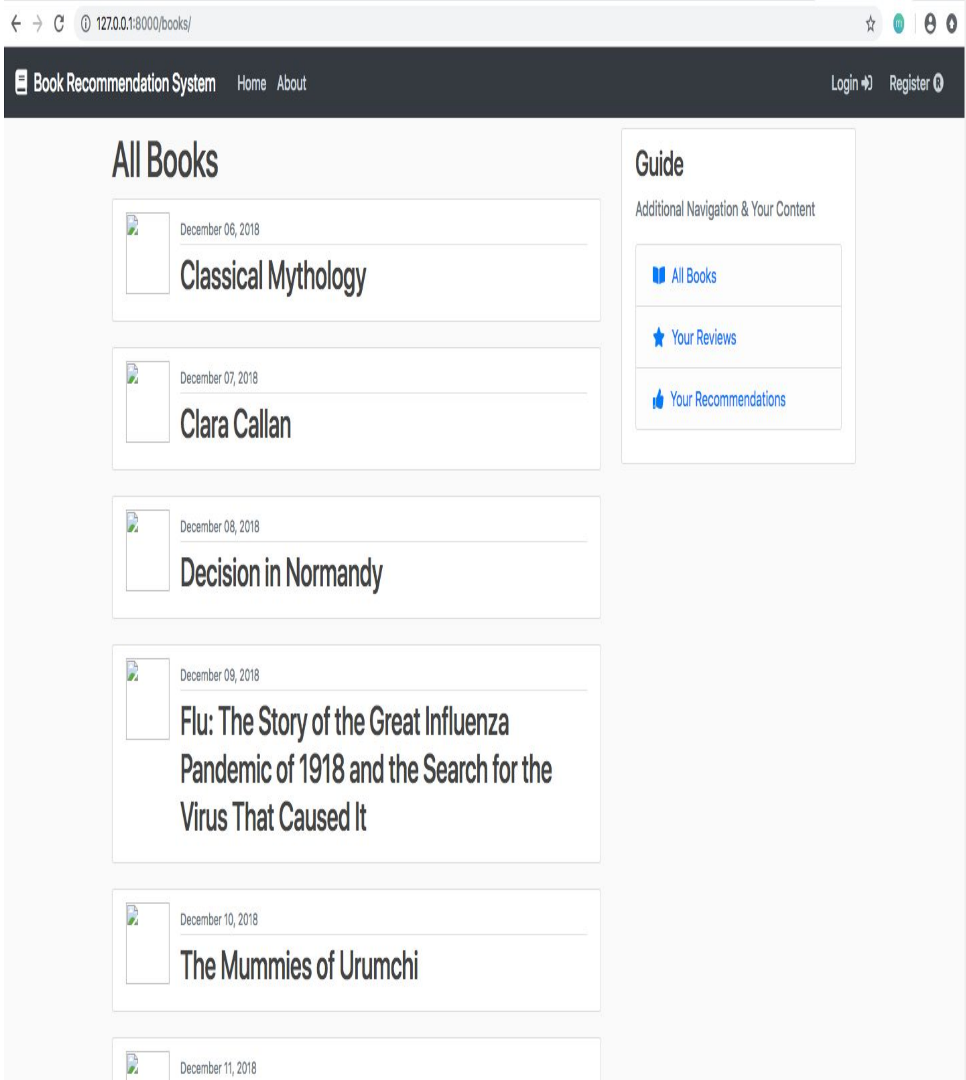
Test Content 2

Guide

Additional Navigation & Your Content

- All Books
- Your Reviews
- Your Recommendations

All Books



All Books

December 06, 2018

Classical Mythology

December 07, 2018

Clara Callan

December 08, 2018

Decision in Normandy

December 09, 2018

Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It

December 10, 2018

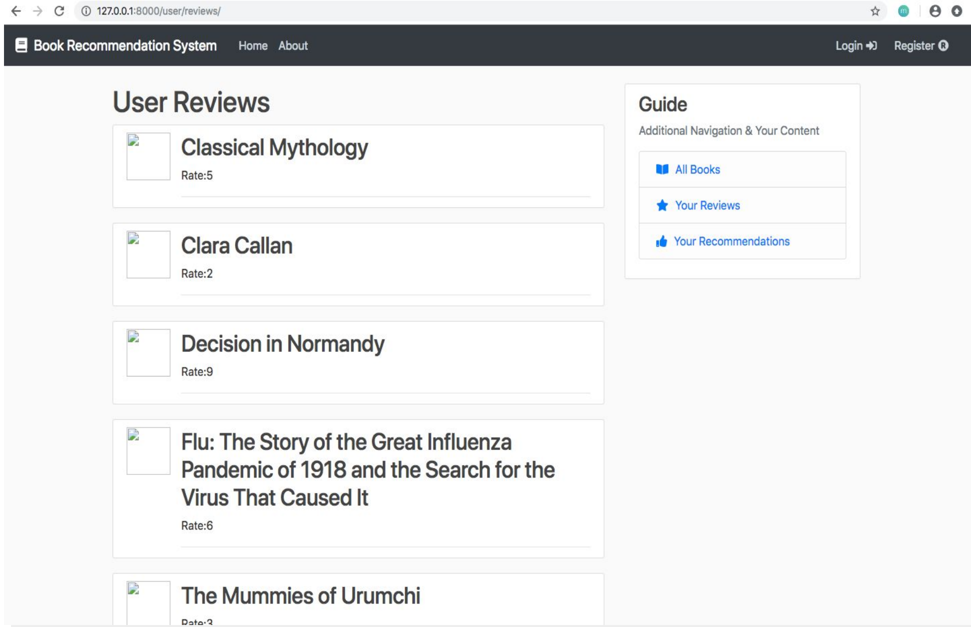
The Mummies of Urumchi

December 11, 2018

Guide

Additional Navigation & Your Content

User Reviews



User Reviews

Classical Mythology

Rate:5

Clara Callan

Rate:2

Decision in Normandy

Rate:9

Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It

Rate:6

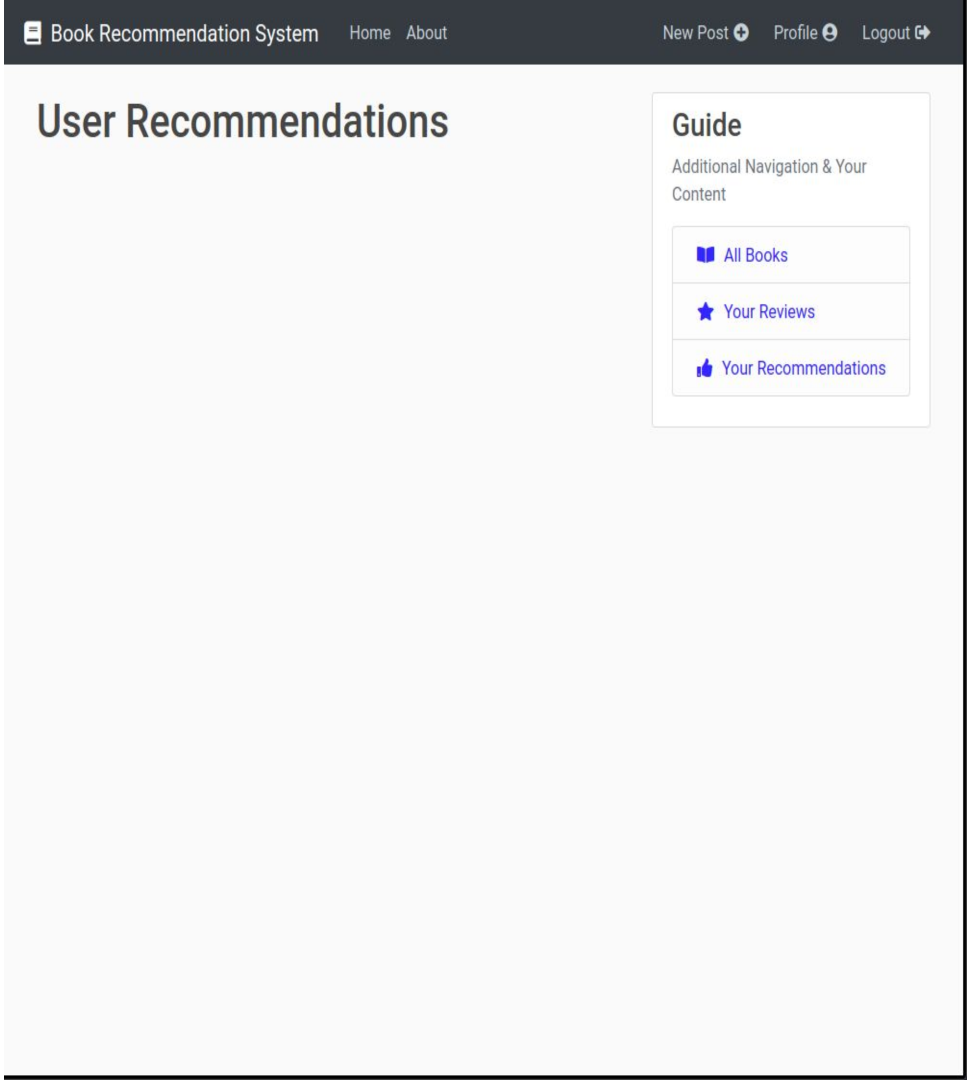
The Mummies of Urumchi

Rate:3

Guide

Additional Navigation & Your Content

User Recommendations



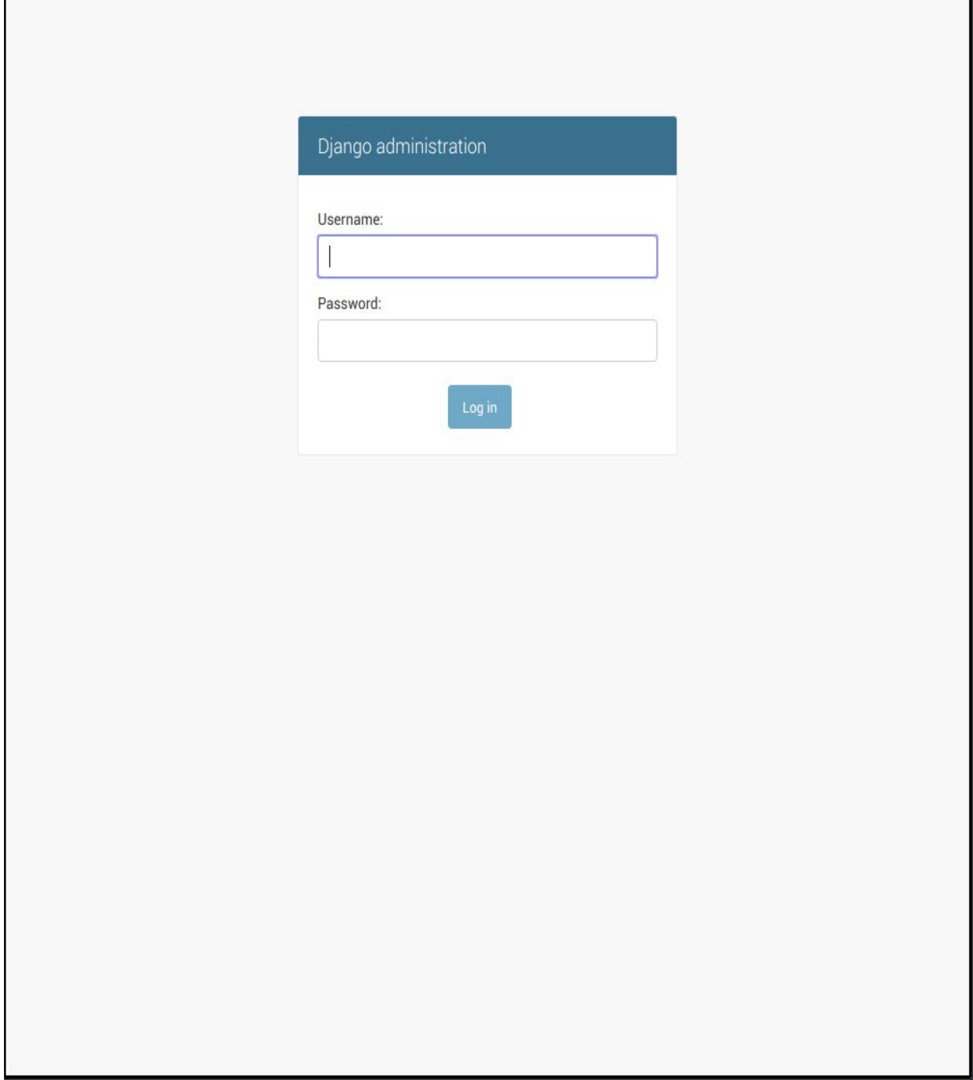
Book Recommendation System Home About New Post Profile Logout

User Recommendations

Guide

Additional Navigation & Your Content

Django Admin

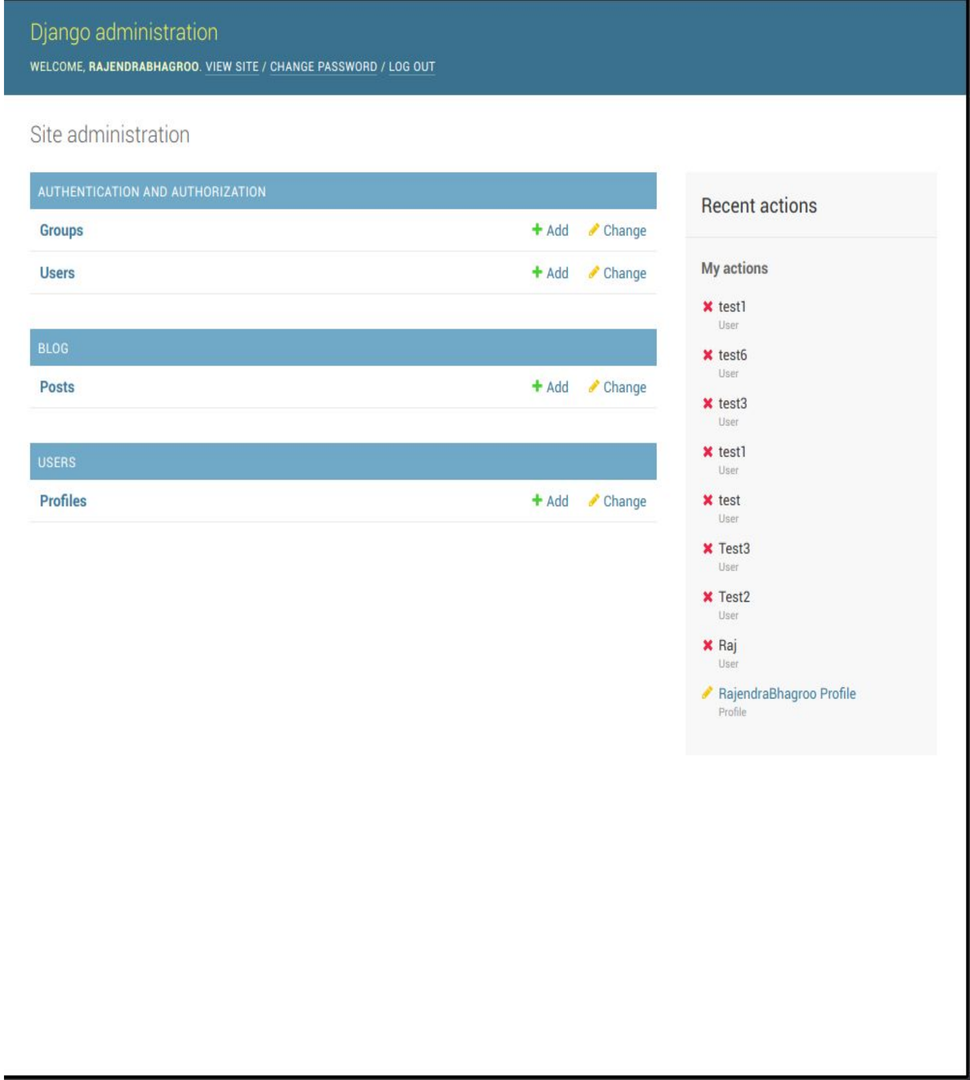


Django administration

Username:

Password:

Admin Homepage



Django administration

WELCOME, **RAJENDRABHAGROO**. VIEW SITE / CHANGE PASSWORD / LOG OUT

Site administration

AUTHENTICATION AND AUTHORIZATION		
Groups	+ Add	✎ Change
Users	+ Add	✎ Change

BLOG		
Posts	+ Add	✎ Change

USERS		
Profiles	+ Add	✎ Change

Recent actions

My actions

✖ test1
User

✖ test6
User

✖ test3
User

✖ test1
User

✖ test
User

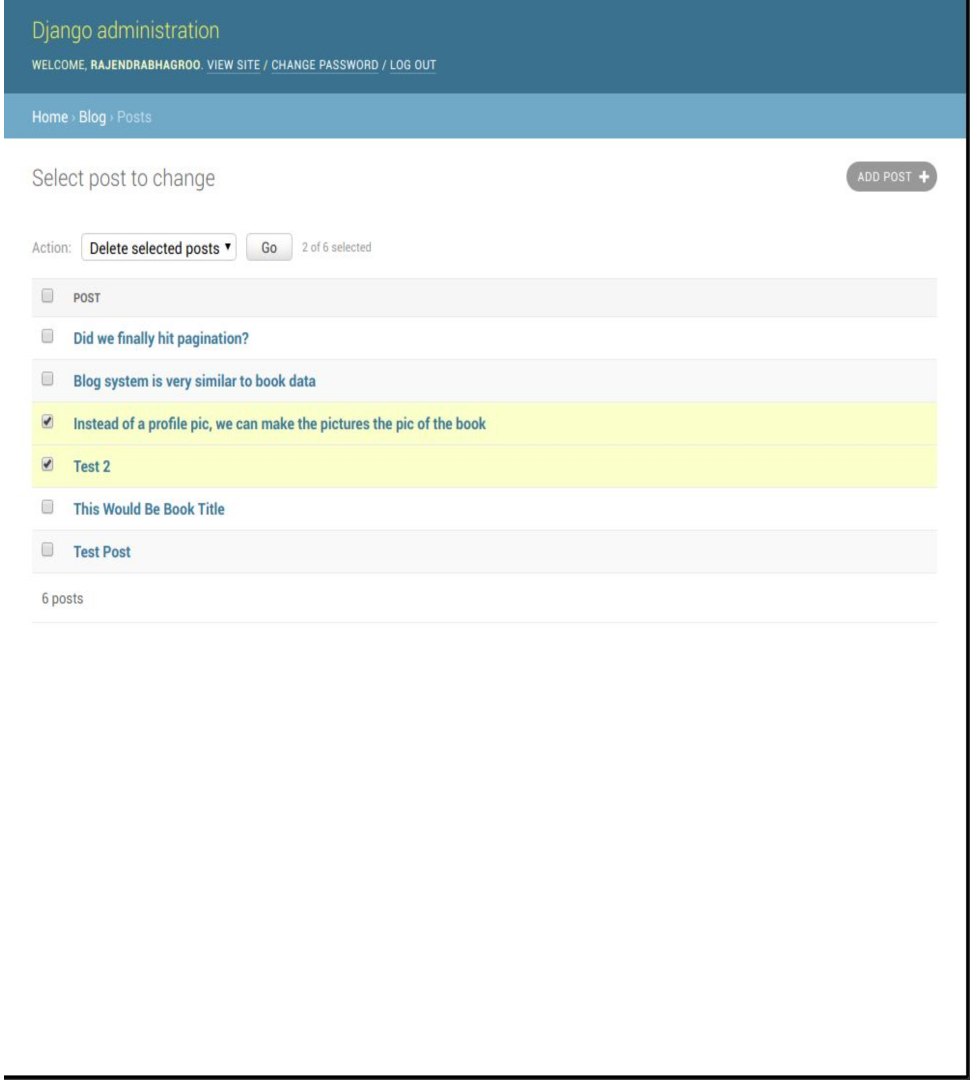
✖ Test3
User

✖ Test2
User

✖ Raj
User

✎ RajendraBhagroo Profile
Profile

Admin Database View



Django administration

WELCOME, **RAJENDRABHAGROO**. VIEW SITE / CHANGE PASSWORD / LOG OUT

Home › Blog › Posts

Select post to change

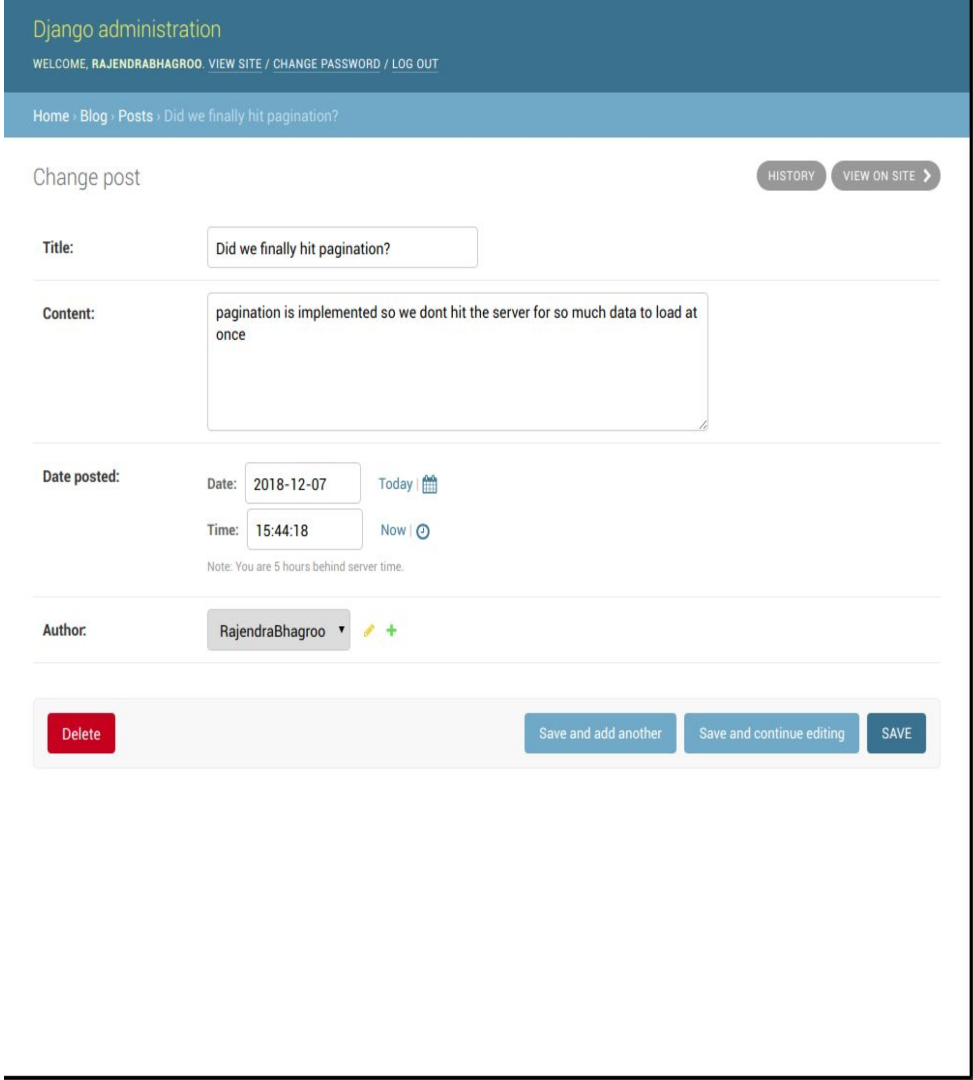
ADD POST ✚

Action: Delete selected posts ▾ Go 2 of 6 selected

☐	POST
☐	Did we finally hit pagination?
☐	Blog system is very similar to book data
☑	Instead of a profile pic, we can make the pictures the pic of the book
☑	Test 2
☐	This Would Be Book Title
☐	Test Post

6 posts

Admin Database Change Data



Django administration
WELCOME, RAJENDRABHAGROO. VIEW SITE / CHANGE PASSWORD / LOG OUT
Home › Blog › Posts › Did we finally hit pagination?
HISTORY
VIEW ON SITE
Today
Now
Delete
Save and add another
Save and continue editing
SAVE

Change post

Title: Did we finally hit pagination?

Content: pagination is implemented so we dont hit the server for so much data to load at once

Date posted:

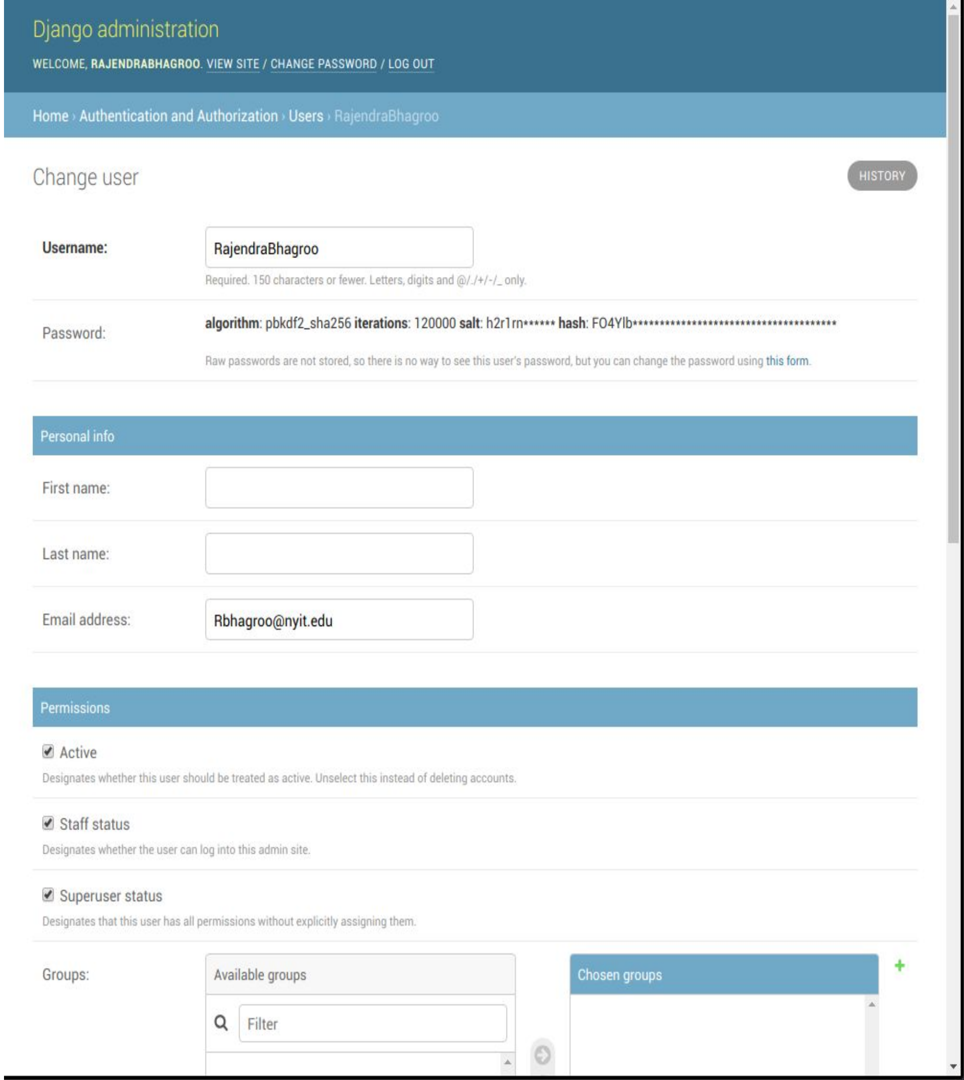
Date: 2018-12-07

Time: 15:44:18

Note: You are 5 hours behind server time.

Author: RajendraBhagroo

Admin
Edit User



Django administration

WELCOME, **RAJENDRABHAGROO**. VIEW SITE / CHANGE PASSWORD / LOG OUT

Home › Authentication and Authorization › Users › RajendraBhagroo

Change user

Username: RajendraBhagroo

Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.

Password: **algorithm**: pbkdf2_sha256 **iterations**: 120000 **salt**: h2r1rn****** **hash**: FO4Ylb**************************************

Raw passwords are not stored, so there is no way to see this user's password, but you can change the password using this form.

Personal info

First name:

Last name:

Email address: Rbhagroo@nyit.edu

Permissions

☑ Active

Designates whether this user should be treated as active. Unselect this instead of deleting accounts.

☑ Staff status

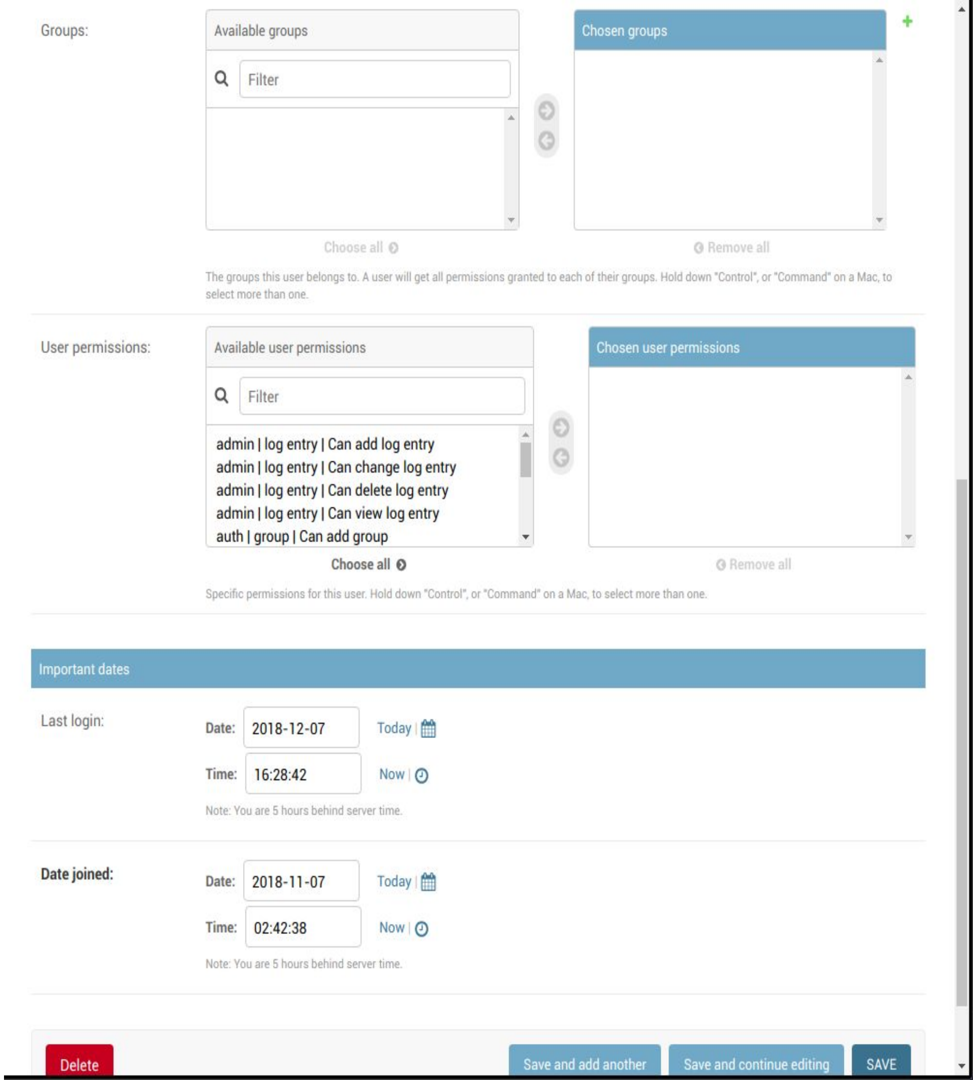
Designates whether the user can log into this admin site.

☑ Superuser status

Designates that this user has all permissions without explicitly assigning them.

Groups:

Admin User Permissions



Groups:

Available groups

Chosen groups

The groups this user belongs to. A user will get all permissions granted to each of their groups. Hold down "Control", or "Command" on a Mac, to select more than one.

User permissions:

Available user permissions

admin | log entry | Can add log entry
admin | log entry | Can change log entry
admin | log entry | Can delete log entry
admin | log entry | Can view log entry
auth | group | Can add group

Chosen user permissions

Specific permissions for this user. Hold down "Control", or "Command" on a Mac, to select more than one.

Important dates

Last login:

Date: 2018-12-07 Today |

Time: 16:28:42 Now |

Note: You are 5 hours behind server time.

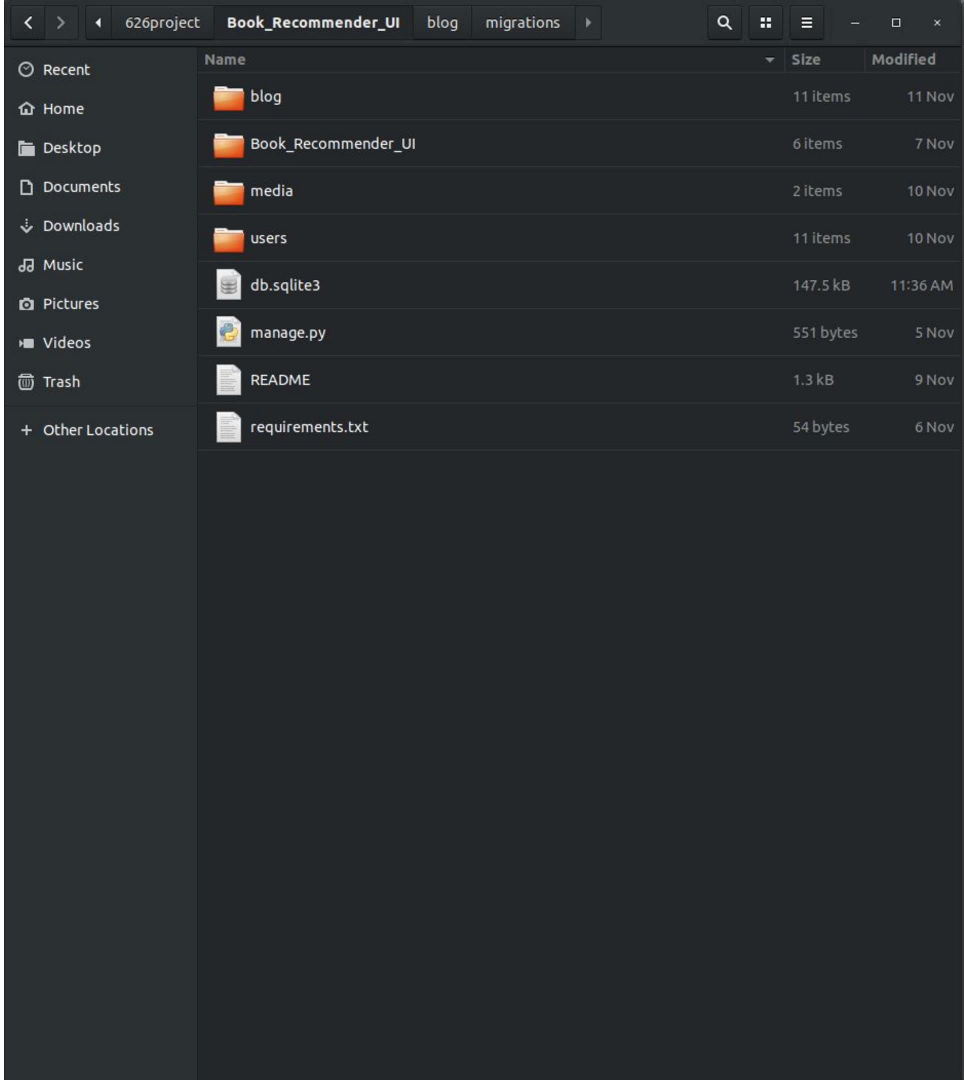
Date joined:

Date: 2018-11-07 Today |

Time: 02:42:38 Now |

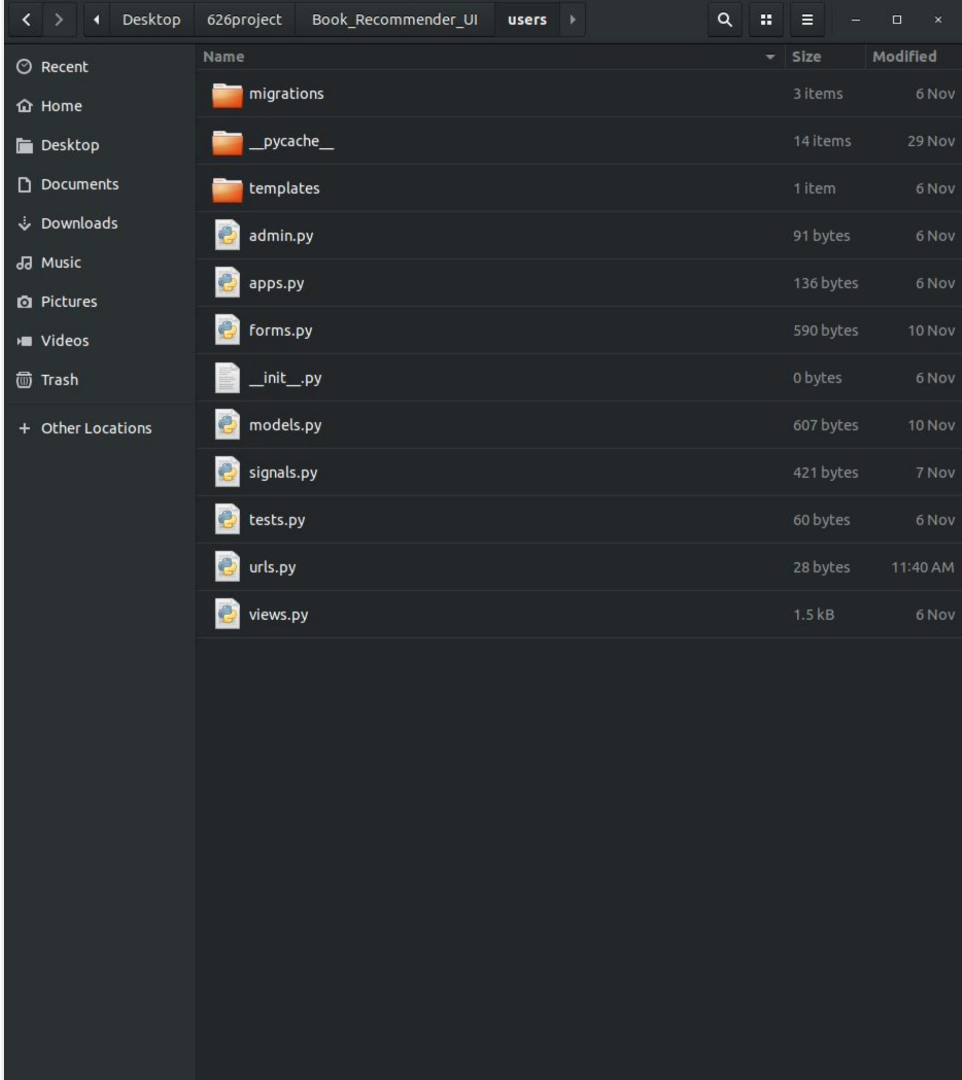
Note: You are 5 hours behind server time.

File Structure UI



Name	Size	Modified
blog	11 items	11 Nov
Book_Recommender_UI	6 items	7 Nov
media	2 items	10 Nov
users	11 items	10 Nov
db.sqlite3	147.5 kB	11:36 AM
manage.py	551 bytes	5 Nov
README	1.3 kB	9 Nov
requirements.txt	54 bytes	6 Nov

File Structure App



Name	Size	Modified
migrations	3 items	6 Nov
__pycache__	14 items	29 Nov
templates	1 item	6 Nov
admin.py	91 bytes	6 Nov
apps.py	136 bytes	6 Nov
forms.py	590 bytes	10 Nov
__init__.py	0 bytes	6 Nov
models.py	607 bytes	10 Nov
signals.py	421 bytes	7 Nov
tests.py	60 bytes	6 Nov
urls.py	28 bytes	11:40 AM
views.py	1.5 kB	6 Nov

Conclusion

Conclusion

- Our book recommendation system will recommend the appropriate books for customers according to their interests and Django will store the recommendations in the customers web profile.
- The system will store the details of the book which users have previously rated and it will use collaborative filtering to recommend new books.
- Collaborative filtering will show a list of all books that are searched based on the content and rating of the book.

Future Work

- In the future, we would like to use clustering algorithms to discover unique groups of users within our application and implement content based filtering to tailor recommendations for all users.
- We would CLEARLY also like to integrate the model into our user interface!