



BookMentor.rs

preporučivač knjiga za studente
Bogdan Marković, Mirko Krstić

Mentor: prof. dr Milan Bjelica

Beograd, Jul 2023.



*Beograd, Mesečniku fakulteta
Beograd, Tehnički fakultet*



Sadržaj

1 Uvod

- ▶ Uvod
- ▶ Data Engineering
- ▶ Web aplikacija
- ▶ Preporučivač
- ▶ Zaključak



Zašto BookMentor

Uvod

- Web aplikacija koja preporučuje knjige studentima Elektrotehničkog fakulteta
- Ideja je nastala kao želja da se na najbolji način izabere sledeća knjiga koju će student pročitati





Izazovi u projektu

Uvod

- Razvijanje projekta od nule
- Nepostojanje opštedostupne baze stručnih knjiga i kurseva fakulteta
- Infrastruktura za aplikaciju
- Izbor algoritma preporučivanja



Sadržaj

2 Data Engineering

- ▶ Uvod
- ▶ Data Engineering
- ▶ Web aplikacija
- ▶ Preporučivač
- ▶ Zaključak



Dobavljanje podataka

Data Engineering

Bilo je potrebno napraviti dve velike baze, i spremiti ih za procesiranje:

- Baza kurseva fakulteta, ručnim labelovanjem
- Baza knjiga, korišćenjem O'Reilly Platform Search API-ja

	Id [PK] integer ↗	Kurs text	o/l "char" ↗	ER boolean ↗	SIMinGod integer ↗	IRMInGod integer ↗	OSMinGod integer ↗	OTMinGod integer ↗	OGMinGod integer ↗	OFMinGod integer ↗	OEMinGod integer ↗	Topics text[]
1	1	Matematika 1	o	true	1	[null]	[null]	[null]	[null]	[null]	[null]	{Algebra:0.3,Linear Algebra:0.3,Calculus:0.3,Math:0.6}
2	2	Osnovi elektrotehnike 1	o	true	[null]	[null]	[null]	[null]	[null]	[null]	[null]	{Microelectronics:0.4,Engineering:0.1,"Electrical Engineering:0.4"}
3	3	Fizika	o	true	1	[null]	[null]	[null]	[null]	[null]	[null]	("Game Physics:0.1",Optics:0.2,Lasers:0.1,Physics:0.8,Science:0.1)



Koriščenje API-ja

Data Engineering

Import books metoda

```
1 def import_books(connection, books_list, page):
2     cur = connection.cursor()
3     iterBook = 0
4     for book in books_list:
5         try:
6             book = process_book(book)
7             if book is None:
8                 continue
9             cur.execute(
10                 "INSERT INTO \"Books\" (isbn, issued, authors, publishers,
11                  title, description, average_rating, popularity, report_score,
12                  cover_url, topic) VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s)",
13                 (book['isbn'], book['issued'], book['authors'], book['
14                  publishers'], book['title'], book['description'],
15                  //...Nastavak na sledecoj strani
```



Korišćenje API-ja

Data Engineering

Import books metoda

```
1          //... Nastavak prethodne strane
2          book['average_rating'], book['popularity'], book['
report_score'], book['cover_url'],
3          book['topics_payload'][0]['name']))
4          iterBook = iterBook + 1
5          except Exception as error:
6              print('ERROR')
7              print(
8                  'Error occurred for the book ISBN:' + book['isbn'] + '
Title: ' + book['title'] + ' at the page ' + str(
9                  page))
10             print(error)
11             continue
12         cur.close()
13         return iterBook
```




Sadržaj

3 Web aplikacija

- ▶ Uvod
- ▶ Data Engineering
- ▶ Web aplikacija
- ▶ Preporučivač
- ▶ Zaključak



Django framework

Web aplikacija

- Django web aplikacija se deli u manje podgrupe - aplikacije
- Svaka aplikacija se bavi svojim poslom tako da medjusobno ukombinovane rade kao celina

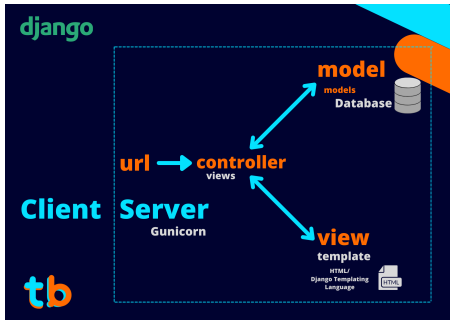
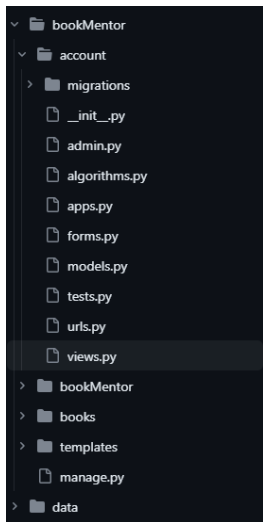


Figure: www.meetgor.com/django-basics-intro/



Struktura web aplikacije

Web aplikacija



- Struktura svake aplikacije je ista
 - postoje `models.py`, `admin.py`, `views.py`, `forms.py`, `urls.py` ...
 - u svakom od fajlova se opisuje svaki deo celog projekta
- pored obaveznih, tu su i fajlovi koji sadrže dodatan kod ili koji predstavljaju satičke datoteke za korišćenje
- `algorithms.py` - u kome se nalaze funkcije za rad sa podacima u procesu preporučivanja
 - ovaj fajl je nestandardan, i bitan samo za aplikaciju `account`



Preporučene knjige

Web aplikacija

[bookMentor.rs](#) [All Books](#) [All Courses](#) [Bogdan's Dashboard](#) [Logout](#)

Recommended books for you

[9781098149468 | The Engineering Executive's Primer](#)

Written by Will Larson

[9781803234199 | Full-Stack Web Development with Go](#)

Written by Nanik Tolaram, Nick Glynn

[9781803247762 | 40 Algorithms Every Programmer Should Know - Second Edition](#)

Written by Imran Ahmad

[9780136823575 | A Tour of C++, 3rd Edition](#)

Written by Bjarne Stroustrup

[9781491910771 | Head First Java, 3rd Edition](#)

Written by Kathy Sierra, Bert Bates, Trisha Gee

[Previous](#) [1 of 3](#) [Next](#)



Sadržaj

4 Preporučivač

- ▶ Uvod
- ▶ Data Engineering
- ▶ Web aplikacija
- ▶ Preporučivač
- ▶ Zaključak



Collaborative filtering algoritam

Preporučivač

- Potrebno je naći korisnike koji su slični - koristimo tabelu tagova

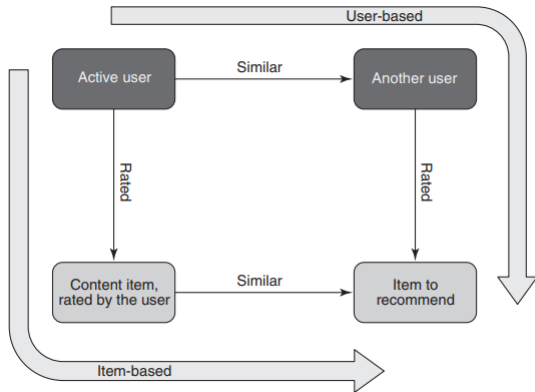


Figure: Kim Falk - Practical Recommender Systems, str. 184



Collaborative filtering algoritam

Preporučivač

- Potrebna je normalizacija feature-a pre pokretanja algoritma
- Poziva se metoda kosinusne sličnosti, iz paketa scikit-learn, pri čemu se vodi računa o stvarnom ID-ju korisnika

$$\text{cosine_similarity}(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}}$$

- Pročitane knjige top N procenata najslićnijih korisnika se dodaju u Rečnik, sa težinskim koeficijentom, kao knjige kandidati



Kombinovanje sa opisom sadržaja

Preporučivač

- Da bi se ublažio problem hladnog starta, i poboljšale preporuke, dodat je algoritam koji koristi opis sadržaja kurseva
- Bira se top N ocenjenih tagova korisnika, a zatim se u prethodni Rečnik dodaju najpopularnije knjige svakog taga, sa svojim težinskim koeficijentom
- Preporučiće se knjige sa najvećim koeficijentom, uz uslov da ih korisnik još nije pročitao



Sadržaj

5 Zaključak

- ▶ Uvod
- ▶ Data Engineering
- ▶ Web aplikacija
- ▶ Preporučivač
- ▶ Zaključak



Potencijalna unapredjenja

Zaključak

- Pipeline za automatsko dodavanje knjiga
- Automatsko ocenjivanje kurseva na osnovu sadržaja
- Poboljšavanje algoritma
- Sreivanje korisničkog interfejsa
- Proširivanje na naučne radove, snimke sa konferencija...



Izvori

Zaključak

- *Kim Falk - Practical Recommender Systems, Manning Publications, 2019.
- *Antonio Mele - Django 4 By Example, Packt Publishing, 2022.



Q&A

Hvala na pažnji!