

Звіт до міні проєкту з ОП 1 семестру 2020

студента 1 курсу КН ФПН УКУ

Пасічника Михайла

Проблема проекту та її вирішення

Зараз у світі більшість людей не дуже люблять читати, оскільки є альтернатива: фільми. Також навіть якщо людина захоче прочитати книжку, їй потрібно буде знайти, що не є простим завданням.

Оскільки при хорошому підборі книжок для користувача, йому дуже ймовірно може сподобатись читання, то мій проєкт очевидно буде популяризувати читання серед людей будь-якого віку, які вже навчилися читати.

Результати розробки моєї програми можуть бути корисними при подальшому вдосконаленні та розробці алгоритмів рекомендацій.

Опис ідеї програми

Я хочу створити сайт, на якому люди могли б, після перегляду фільму, знайти книжку, яку можна прочитати, а після прочитання книжки знову ж таки користувачеві видались назви фільмів, які могли б бути йому цікаві, що буде в якості приза і заохочуватиме користувачів до читання. Також на ньому буде рейтингова система користувачі, через що внаслідок конкуренції буде збільшуватись бажання користувачів читати.

Тема дослідження

Оскільки головною задачею мого проєкту є саме складання алгоритмів рекомендацій, то я хочу саме це і зробити. Моя програма буде шукати топ книг, які можуть сподобатись користувачеві в залежності від фільму, який він подивився. Моя програма буде давати кожній книзі оцінку схожості до даного фільму, і ця оцінка кожен раз буде додаватись до попередньої, що буде створювати накопичувальну систему уподобань користувача і рекомендації ставатимуть все кращими.

Загальний опис програми

У мене в програмі є 4 основні модулі:

1. books_info_read.py - модуль який працює з базами даних книжок, опрацьовує їх за допомогою бібліотек: pandas, csv, re, json. Результуюча база даних книжок вже використовується в основному модулі алгоритму пошуку рекомендацій.

Я використав бази даних з наступних ресурсів:

- <https://www.kaggle.com/zygmunt/goodbooks-10k>
- <https://www.kaggle.com/iamhungundji/book-summary-genre-prediction>

Звідси я використовував дані про назви книжок, їх авторів, дати публікації, рейтинг, жанри, опис книжок та кількість сторінок.

2. `films_info_read.py` - модуль який працює з базами даних фільмів, опрацьовує їх за допомогою бібліотек: `pandas`, `ast`. Результуюча база даних фільмів вже використовується в основному модулі алгоритму пошуку рекомендацій.

Я використав бази даних з наступних ресурсів:

- https://www.kaggle.com/rounakbanik/the-movies-dataset?select=movies_metadata.csv

Звідси я використовував дані про назви фільмів, їх номер IMDB, жанри, опис фільмів, дату релізу, назви, рейтинг та список ключових слів про фільми.

3. `crate_recommendations.py` - модуль, який власне шукає рекомендації книжок, через дані з баз даних оброблених в попередніх двох модулях. Тут використовуються бібліотеки: `pandas` та `sklearn`. Алгоритм пошуку відбувається через співставлення жанрів, пошук ключових слів в описах книжок, а також на співставленні описів книжки та фільму за допомогою бібліотеки `sklearn`.
4. `main.py` - основний модул, який об'єднує всі попередні в робочу програму. При запуску цього модуля в терміналі слідуючи вказівкам можна користуватися програмою за шукати рекомендації книг, в залежності від переглянутого фільму.

У всіх модулях використовується бібліотека `tqdm` для відображення стану роботи програми. При розробці усіх модулів я активно користувався наступним сайтами для пошуку необхідної моні інформації щодо написання коду:

- <https://stackoverflow.com/>
- <https://pandas.pydata.org/>
- <https://towardsdatascience.com/building-a-content-based-book-recommendation-engine-9fd4d57a4d>

Усі дані які я використовую є у вільному доступі.

Отримані результати

```
➤ recommendations git:(master) ✖ python main.py
reading books info
16559it [00:00, 32138.41it/s]
100%|██████████| 16559/16559 [00:00<00:00, 542834.32it/s]
100%|██████████| 10957/10957 [00:03<00:00, 3477.24it/s]
complete reading books info
links.csv
reading films_info
movies_metadata.csv
complete reading films_info
-----
__init__.py      Hi!
You see the film and won't read some book in same theme?
Print 'y' if yes and something else if not.
y
books_info_read.py
crate_recommendations.py
sdskldsk films_info.csv
Name is wrong! Try again
jumanji
Print number of film which exactly you watched:
1 - Jumanji
Write number: 1
Wait a little bit. It will cost it :)
100%|██████████| 10834/10834 [00:18<00:00, 576.98it/s]
100%|██████████| 1979/1979 [00:00<00:00, 13579.53it/s]
100%|██████████| 10834/10834 [00:00<00:00, 69358.45it/s]
Hear are recommendations for you:
Book #1: test2.py
Title: Someone Comes to Town, Someone Leaves Town
Author: Cory Doctorow
Genre: Fantasy Speculative fiction
Rating: 3.91/5
Publication date: 10/15/2013
num_pages: 255
Description: the story mainly takes place in two ontario localities in flashbacks the main character usually but not always called alan he appears to have been alphabetized rather than named and will answer to any masculine name beginning with a and his brothers also alphabetized grow up outside of the remote town of kapuskasing the novel opens with alans purchase of a home in the kensington market neighborhood of modern day toronto there are two main plotlines alan befriends kurt a thirtysomething punk who operates a dumpster diving operation kurt uses computer components that he retrieves from the trash and turns them into functioning wi fi access points kurts goal is to blanket the entire neighborhood with free and secure internet access by attaching his access points to buildings with the permission of their owners kurts plan doesnt really get off the ground until he forms a partnership with alan who puts a more professional face on the operation and sweet talks many local owners into allowing the access points to use their space and a small amount of their electricity the second plotline features fantasy elements unbeknownst to most of the other characters alan and his brothers are not quite human their father is a mountain and their mother is a washing machine alans eldest brother can see the future his second eldest is an island his third eldest is undead and his three youngest brothers are a set of russian nesting dolls alan is the most normal seeming of his family outwardly he looks human but he heals at an incredible rate and if part of him is cut off it will grow back and the cut off part can be made to form a new copy of him much like an earthworm does another plot strand concerns alans neighbours a household of students and artists which includes mimi a troubled young woman who like alan is not quite human born with wings on her back and no family history she lives with her abusive boyfriend krishna a musician bartender who can spot beings like alan and his family and hates them krishna amputates mimis wings every three months she stays with him because she believes hes the only one willing and able to make her normal
Print number of film
- Alanji
(its number) 1
As many book you will read,
as good recommendation you will have!
Will you find next book!
Print 'y' if yes and something else if not.
y
Ok! Have nice day!
Title: Someone Comes to Town,
Author: Cory Doctorow
Genre: Fantasy Speculative fi
num_pages: 255
```

```
Rating: 3.91/5
Publication date: 10/15/2013
num_pages: 255
Description: the story mainly takes place in two ontario localities in flashbacks the main character usually but not always called alan he appears to have been alphabetized rather than named and will answer to any masculine name beginning with a and his brothers also alphabetized grow up outside of the remote town of kapuskasing the novel opens with alans purchase of a home in the kensington market neighborhood of modern day toronto there are two main plotlines alan befriends kurt a thirtysomething punk who operates a dumpster diving operation kurt uses computer components that he retrieves from the trash and turns them into functioning wi fi access points kurts goal is to blanket the entire neighborhood with free and secure internet access by attaching his access points to buildings with the permission of their owners kurts plan doesnt really get off the ground until he forms a partnership with alan who puts a more professional face on the operation and sweet talks many local owners into allowing the access points to use their space and a small amount of their electricity the second plotline features fantasy elements unbeknownst to most of the other characters alan and his brothers are not quite human their father is a mountain and their mother is a washing machine alans eldest brother can see the future his second eldest is an island his third eldest is undead and his three youngest brothers are a set of russian nesting dolls alan is the most normal seeming of his family outwardly he looks human but he heals at an incredible rate and if part of him is cut off it will grow back and the cut off part can be made to form a new copy of him much like an earthworm does another plot strand concerns alans neighbours a household of students and artists which includes mimi a troubled young woman who like alan is not quite human born with wings on her back and no family history she lives with her abusive boyfriend krishna a musician bartender who can spot beings like alan and his family and hates them krishna amputates mimis wings every three months she stays with him because she believes hes the only one willing and able to make her normal
Print number of film
- Alanji
(its number) 1
As many book you will read,
as good recommendation you will have!
Will you find next book!
Print 'y' if yes and something else if not.
y
Ok! Have nice day!
Title: Someone Comes to Town,
Author: Cory Doctorow
Genre: Fantasy Speculative fi
num_pages: 255
```

Результатом виконання мого міні-проекту є програма, через яку в терміналі можна отримувати рекомендації книжок на основі переглянутих фільмів. Щоб скористатися моєю програмою потрібно устанувати дистрибутив і в відповідну папку проекту розархівувати наступний архів з необхідними базами даних.

<https://drive.google.com/file/d/1jDdl2NNBqVVLmmCVKODen2VIEKmQ8nXd/view?usp=sharing>

Як показано на прикладі, дотримуючись інструкцій в терміналі користувач з легкістю може зрозуміти, як користуватись програмою і, відповідно, отримати хороші рекомендації книжок.

На мою думку моя програма може повністю виправдати мої сподівання від неї та максимально популяризувати читання, оскільки якщо б існувала така веб-система, де були б хороші рекомендації та можливість змагатися з іншими в такій сфері, як читання, то вона б розвивалася навпаки, а не занепадала. Проте до моєї кінцевої ідеї ще дуже далеко, оскільки щоб створити такий сайт та популяризувати його потрібно дуже багато інформаційних та матеріальних ресурсів.