Developer Documentation for Sentiment-Enhanced Book Recommendation System

Group 2: Andrew Stokey, Eli Kenemore, Jack Dodge

December 2nd, 2024

Contents

1 System Overview

All datasets and Python scripts are available in the GitHub repository:
https://github.com/stokeya/Books-Recommendation-System-Part2-.
The project is structured as follows:

- `data_cleanup.py`: Cleans and merges datasets.
- `test_data.py`: Splits the dataset into training and testing sets.
- `bookrecsystemcleaned.py`: Implements the recommendation logic.

2 Development Environment

Developers can use either VS Code or PyCharm to run and modify the scripts. Ensure Python 3.7 or higher is installed along with all required libraries.

3 Key Components

3.1 `data_cleanup.py`

This script processes two raw datasets by cleaning null values, removing unnecessary columns, and merging the data.

```
merged_df = pd.merge(ratings_df, data_df)
merged_df.dropna(subset=['Title'], inplace=True)
merged_df.to_csv('Resources/complete_cleaned_data.csv')
```

Listing 1: Example: Cleaning and Merging Datasets

3.2 `test_data.py`

This script splits the cleaned dataset into training and testing subsets.

```
test_data = unique_titles.sample(n=100000, random_state=42)
train_data = data_df[~data_df["Title"].isin(test_data["Title"])]
test_data.to_csv("test_dataset.csv")
train_data.to_csv("train_dataset.csv")
```

Listing 2: Example: Dataset Splitting

3.3 `bookrecsystemcleaned.py`

This script performs sentiment analysis using TextBlob and implements a kNN-based recommendation system.

```
def analyze_sentiment(text):
    blob = TextBlob(text)
    return round((blob.sentiment.polarity + 1) * 2 + 1)
```

Listing 3: Sentiment Analysis

4 Future Development

- Incorporate advanced NLP models like transformers for sentiment analysis.
- Build a GUI for better user experience.
- Add hierarchical filtering for categories and genres.