Book Reccomendations.ipynb

2023/12/07

A Birds Eye View

Using: Pandas + Web scraping + Weighed cosine similarity ranking

Developed a book recommendation system incorporating web scraping for Wikipedia book descriptions. The system tailors recommended books based on location, date, subject, author, and category. This project showcases knowledge in data preprocessing, recommendation algorithms, and web scraping.

1.1 Loading Data

row_id	Title	Vendor	Tags	Description	Location	Category	Date	Download Link	WikiLink
0	Bushido, the Soul of Japan	Inazo Nitobe	19th century	Bushido: The Soul of Japan is a book written b...	Japanese	Philosophy	1899	https://www.gutenberg.org/cache/epub/12096/pg1...	https://en.wikipedia.org/wiki/Bushido:_The_Sou...
1	The Beautiful and Damned	F. Scott Fitzgerald	20th century	The Beautiful and Damned is a tragic novel by ...	American	Fiction	1922	https://www.gutenberg.org/cache/epub/9830/pg98...	https://en.wikipedia.org/wiki/The_Beautiful_an...

1.2 Cleaning Data

- Used dropna to eliminate null values (**DropNA**) in critical fields–**WikiLink, tags, title, author, date, location.**
- Vendor column renamed to Author for clarity.
- Update the row_id values to represent the new row position.

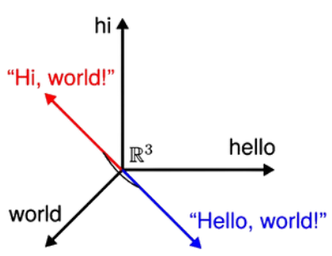
1.3 Feature Engineering

Used BeautifulSoup to extract a brief snippet from the initial paragraph of the provided link, enhancing the dataset's informativeness for improved book recommendations.

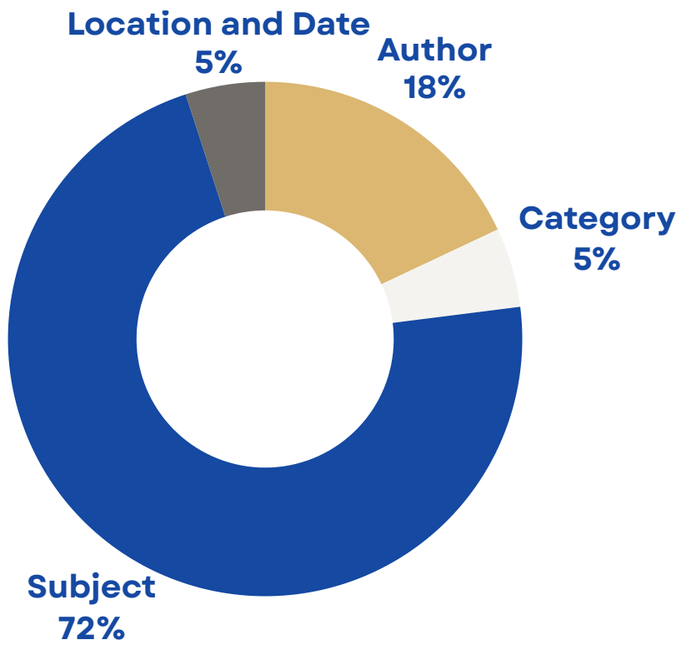
Subject example: War and Peace a is a literary work by Russian author Leo Tolstoy Set during the Napoleonic Wars the work mixes fictional narrative with chapters discussing history and philosophy...

1.4 Recommendation Ranking

1. Transformed text data into word count matrices for Subject, Vendor, Category, Location, and Date.
2. Computed cosine similarity for each matrix.
3. Matrices were weighted to make a combined score (Vendor 18%, Location/Date 5%, Subject 72%, Category 5%).
4. Scores are sorted and then ranked to extract the book recommendations.



Cosine Similarity



Thomas Cook