Datasheet for LibraryThingFull

Motivation for Dataset Creation

Why was the dataset created? This dataset was created to support investigation into films that were made as adaptations of novels.
What (other) tasks could the dataset be used for? We do not recommend using this dataset for anything else.
Has the dataset been used for any tasks already? This datasheet was used for a school project. Please see https://github.com/MRWilliamsGit/BooksToMoviesML
Who funded the creation of the dataset? This dataset was created for free from free resources.
Any other comments?

Dataset Composition

What are the instances? Each observation is of a search done on libraryThing.com. Each search term was generated by an instance in a list of films tagged as novel adaptations on IMDb.com. If there is 'adaptation of' and 'Link2' data, the x features relate to that adaptation. If there is not, the x features relate to the Searchterm and Link1.
Are relationships between instances made explicit in the data? N/A
How many instances of each type are there? There are 1160 observations.
What data does each instance consist of? Each observation includes the original search term, the link it generated, the title of the work cited as its source material, the link for the source material, and then publisher and author information, whether the work was part of a series, its rating, number of listed characters, and number of awards attributed to it.
Is everything included or does the data rely on external resources? This data is as it was on LibraryThing.com in November of 2021.
Are there recommended data splits or evaluation measures? N/A
What experiments were initially run on this dataset? This datasheet was used for a school project. Please see https://github.com/MRWilliamsGit/BooksToMoviesML
Any other comments?

Data Collection Process

How was the data collected? This data was collected using a scraping script using a list of movies generated by another scraping script.
Who was involved in the data collection process? This data was collected by a master's student, for free.
Over what time-frame was the data collected? This data was collected in several batches in November of 2021.
How was the data associated with each instance acquired? This information is all crowd-generated from public knowledge.
Does the dataset contain all possible instances? No, the data collected is only of specific returns from specific searches on this website.
If the dataset is a sample, then what is the population? At the time of this writing in Nov, 2021, LibraryThing has information on over 155 million books, according to wikipedia.com.
Is there information missing from the dataset and why? There is a lot of information missing simply due to the crow-sourced nature of the data.
Are there any known errors, sources of noise, or redundancies in the data? N/A
Any other comments?

Datasheet for LibraryThingFull

Data Preprocessing

What preprocessing/cleaning was done? No preprocessing was done on this dataset after it was pulled.
Was the "raw" data saved in addition to the preprocessed/cleaned data? N/A
Is the preprocessing software available? N/A
Does this dataset collection/processing procedure achieve the motivation for creating the dataset stated in the first section of this datasheet? N/A
Any other comments?

Dataset Distribution

How is the dataset distributed? This dataset is hosted on Github at https://github.com/MRWilliamsGit/BooksToMoviesML
When will the dataset be released/first distributed? November 2021
What license (if any) is it distributed under? N/A
Are there any fees or access/export restrictions? No
Any other comments?

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset?
This dataset is only hosted on https://github.com/MRWilliamsGit/BooksToMoviesML
Will the dataset be updated? This dataset will most likely not be updated.
If the dataset becomes obsolete how will this be communicated? This dataset will always be representative of a November 2021 scrape of specific search terms.
Is there a repository to link to any/all papers/systems that use this dataset? https://github.com/MRWilliamsGit/BooksToMoviesML
If others want to extend/augment/build on this dataset, is there a mechanism for them to do so? Not at this time
Any other comments?

Legal & Ethical Considerations

If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection? N/A
If it relates to other ethically protected subjects, have appropriate obligations been met? N/A
If it relates to people, were there any ethical review applications/reviews/approvals? N/A
If it relates to people, were they told what the dataset would be used for and did they consent? N/A
If it relates to people, could this dataset expose people to harm or legal action? N/A
If it relates to people, does it unfairly advantage or disadvantage a particular social group? N/A
If it relates to people, were they provided with privacy guarantees? N/A
Does the dataset comply with the EU General Data Protection Regulation (GDPR)? N/A
Does the dataset contain information that might be considered sensitive or confidential? N/A
Does the dataset contain information that might be considered inappropriate or offensive? N/A
Any other comments?