

Yasaman Sabersheikh

yasaman2@illinois.edu

CS410 – Fall 2021 – Text Transformation Systems

Final Project – 1 person team

Goodreads Scraper

OVERVIEW

Goodreads is a free website on books that allows individuals to freely search for books and reviews.

Users can sign up and generate library and reading lists. <https://www.goodreads.com/>

Finding books that are award winners are not easy on the website. Goodreads has a link for Reader's Choice awards but other than that it is hard to find other literary awards such as The Man Booker Prize or Giller Prize.

In this project I intend to write a web scraper to find the books that award winners or have a high rating. The result can be used for book recommendation, book clubs or an annual blog post.

GOALS

1. Create a scraper to scrape the data in Goodreads website
2. Can categorize books under the same reward category by year
3. Add the ratings to each book in rewards categories as well
4. Create a Category of highest rating books by year. e.g., 2021 rating 4 and above
5. Creating an HTML page to present the data

SPECIFICATIONS

- Python programming language for scripting
- Selenium to retrieve the Html source code
- BeautifulSoup to extract book-data from the Html source code