Report on progress done on the Goodreads Plus project as of April 23rd and the tasks left to do

Install the extension
Source code
Video demo
Project log

Problem

- I love listening to audiobooks.
- Goodreads is the most popular “book social media” out there.
    - Allows tracking read dates and rating books
- As a computer scientist, I also want to data mine my hobby to gain insights.
- Unfortunately, Goodreads is quite stagnant and lacks useful data export feature.
    - They had public APIs, but those have been removed a few years ago.
    - They have an RSS feed and a takeout option, but those provide limited data and aren’t convenient to use

Solution

A Google Chrome extension that:

- Adds an easy data export feature to Goodreads
- Adds powerful plotting functionality
- Adds comprehensive book search (Goodreads’ search is infamously bad)

Worth mentioning

As part of this project, I wrote a production-ready TypeScript implementation of the Porter-stemmer algorithm and published it on NPM. There have been close to 200 downloads of the package within the first week.

What is more, I reached out to Martin Porter about adding my implementation to the home page of the Porter-stemmer algorithm, and it is now visible there (the last entry in the table) - https://tartarus.org/martin/PorterStemmer/

Completed tasks

- Ability to export a library from Goodreads
    - When exporting, also retrieve the user’s review, rating, and read dates
    - Correctly resolve the page count for audiobooks

        - Handle errors and interruptions during the process
        - Show a progress report
- Charts for visualizing the data
    - Compare reading rates between years
    - Show insights about your favorite books
- Search
    - Porter-stemmer algorithm
    - TF-IDF based search (using inverse index) on multiple metadata fields of a book with different weights for different fields
- Analysis tools
    - min, max, average book length year to year
    - reading pace indicators
- Advanced table features
    - Sorting
    - Filtering
    - Column reordering
    - Changing column visibility
    - Remember the table state between sessions
- Publish to Google Chrome WebStore
    - Provide documentation
    - Provide screenshots
    - Record a demo video
    - Write a privacy policy
    - Submit for review to Chrome WebStore

Remainings tasks

There is a nice-to-have feature that is low priority but high effort: exporting not just the reading dates for books, but the entire timeline.
This includes:

- When you added the book to the “want to read” shelf
- When you started reading the book
- Any progress reports or notes you took during the reading process
- When you finished reading the book
- Likes and responses to your review

While these features won’t be interesting to all but most advanced users, they are not available for export from Goodreads by any means so the extension can provide a useful feature.
Besides the complexity of scraping this content from the interface, there is also a difficulty in presenting time series data in a tabular layout.
For these reasons, the implementation of this feature was deferred.

Code

Tech Stack

- JavaScript
- TypeScript
- React
- Tailwind CSS
- PrimeReact
- ChartJS
- Jest (for Unit test)

Usage of third-party libraries

PrimeReact was used for displaying the main table, the pop-over, and the tabs. This allowed them to quickly remove the responsibility for the undifferentiated common lifting and focus on the more interesting things. Still, I customized these components and improved styling.

Additionally, ChartJS was used to render the charts, as rendering interactive charts without a library in a performant way is very tricky and a bad use of time when there are already a lot of great solutions.

Besides those, 100% of the code was written by me.

Notably, the code for the potter-stemmer algorithm was implemented from scratch based on the original paper, and I used their example files (with 18k words) to verify the correctness of the implementation.

The code for the TF-IDF-based search was written from scratch, although it was guided by the material covered in class.

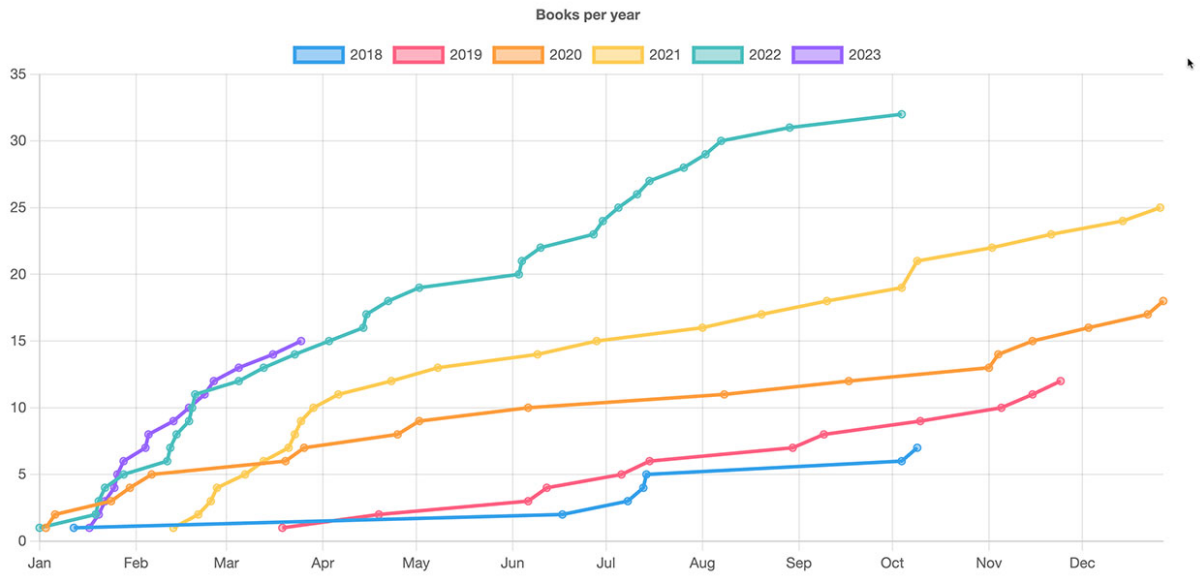
Screenshots

Can compare reading rate between the years

Goodreads Stats

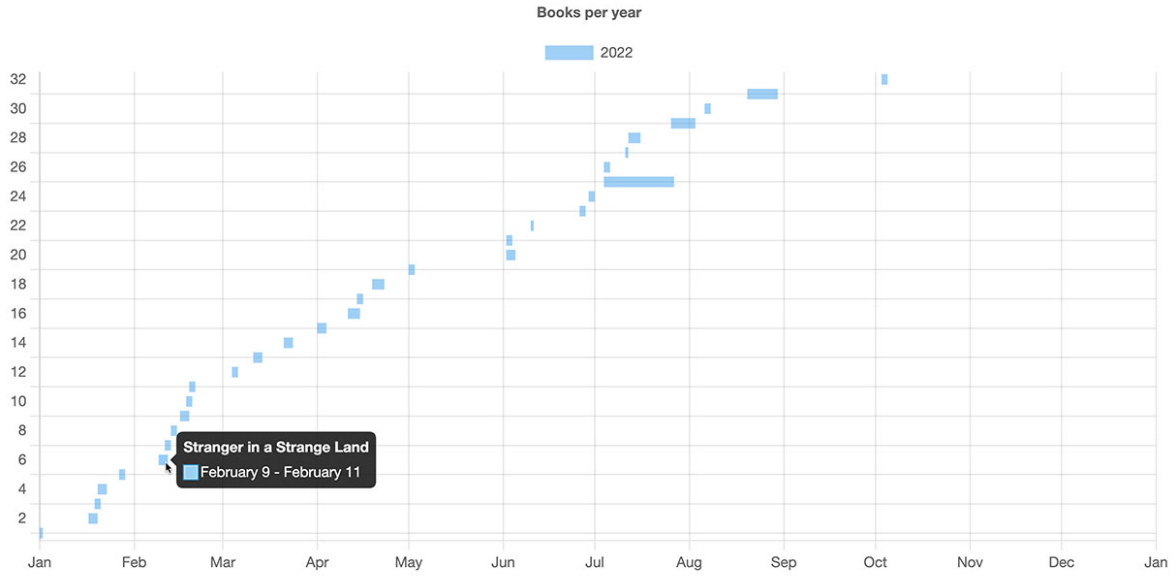
Last updated:last week

Stats | All books | Search



Can see how long it took to read each book

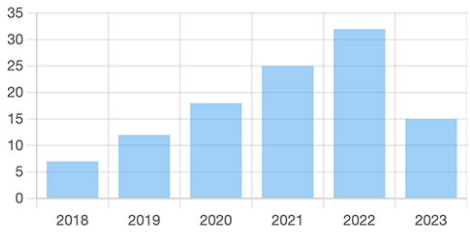
Year 2022



Pages per year

Can get an extensive breakdown and insights

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

	2018	2019	2020	2021	2022	2023
Books						
Pages	2812	4541	8219	9192	12830	5558
Books per day	0.02	0.03	0.05	0.07	0.09	0.04
Pages per day	7.7	12.44	22.46	25.18	35.15	15.23
Books per day (excluding blank days)	0.13	0.19	0.26	0.33	0.34	0.5
Pages per day (excluding blank days)	51.13	72.08	119.12	120.95	136.49	185.27
Average book length	401.71	378.42	456.61	367.68	400.94	370.53
Longest book	384	559	193	559	652	372
Shortest book	364	364	352	364	311	480

Year 2023

Can search and filter your entire library

Goodreads Stats

Last updated:last week

Stats All books Search

Max's bookshelf: favorites

Columns

Image	Title	Author	Page count	Average rating	Rating	Read times
	Elon Musk: Tesla, SpaceX, and the Quest for a Fantastic Future	Ashlee Vance	392	4	5	3/13/2023 - 3/16/2023 10/9/2019 - 10/10/2019 6/12/2018 - 6/17/2018 1/4/2018 - 1/12/2018
	Replay	Ken Grimwood	311	4	5	2/26/2023 - 2/26/2023 12/29/2021 - 1/1/2022
	Warcross (Warcross, #1)	Marie Lu	366	4	5	2/18/2023 - 2/18/2023 12/24/2021 - 12/26/2021

Can download entire library or a single shelf