**About Goodreads**

**Ref:** [**https://en.wikipedia.org/wiki/Goodreads**](https://en.wikipedia.org/wiki/Goodreads)

Goodreads is an American social cataloguing website and a subsidiary of Amazon that allows individuals to search its database of books, annotations, quotes, and reviews. Users can sign up and register books to generate library catalogues and reading lists. They can also create their own groups of book suggestions, surveys, polls, blogs, and discussions. The website's offices are located in San Francisco.

Goodreads was founded in December 2006 and launched in January 2007 by Otis Chandler and Elizabeth Khuri Chandler. In December 2007, the site had 650,000 members and 10,000,000 books had been added. By July 2012, the site reported 10 million members, 20 million monthly visits, and thirty employees. On March 28, 2013, Amazon announced its acquisition of Goodreads, and by July 23, 2013, Goodreads announced their user base had grown to 20 million members.

By July 2019, the site had 90 million members.

**Goodreads Initial Subscription**

**Ref:** [**https://www.goodreads.com/**](https://www.goodreads.com/)

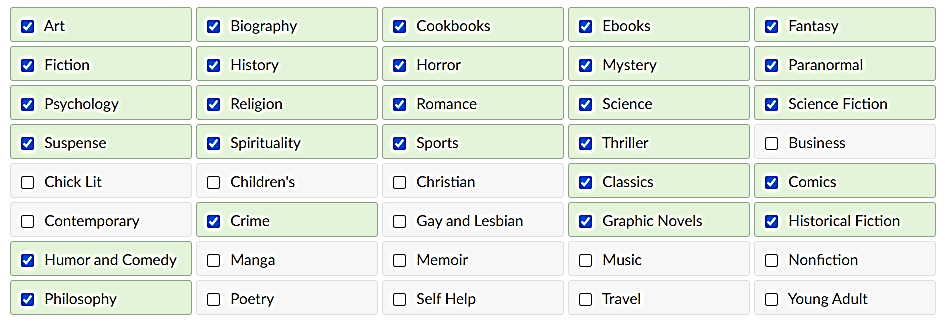
The initial genres offered by Goodreads are as follows.

* Top of Form

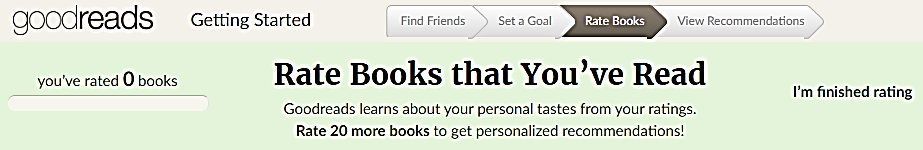
Bottom of Form

* Art
* Biography
* Business
* Chick Lit
* Children's
* Christian
* Classics
* Comics
* Contemporary
* Cookbooks
* Crime
* EBooks
* Fantasy
* Fiction
* Gay and Lesbian
* Graphic Novels
* Historical Fiction
* History
* Horror
* Humour and Comedy
* Manga
* Memoir
* Music
* Mystery
* Nonfiction
* Paranormal
* Philosophy
* Poetry
* Psychology
* Religion
* Romance
* Science
* Science Fiction
* Self Help
* Suspense
* Spirituality
* Sports
* Thriller
* Travel
* Young Adult

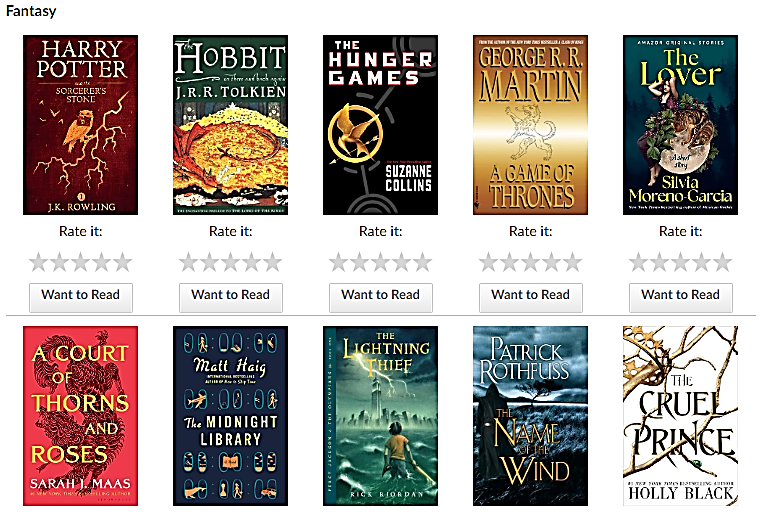
**My selection:**

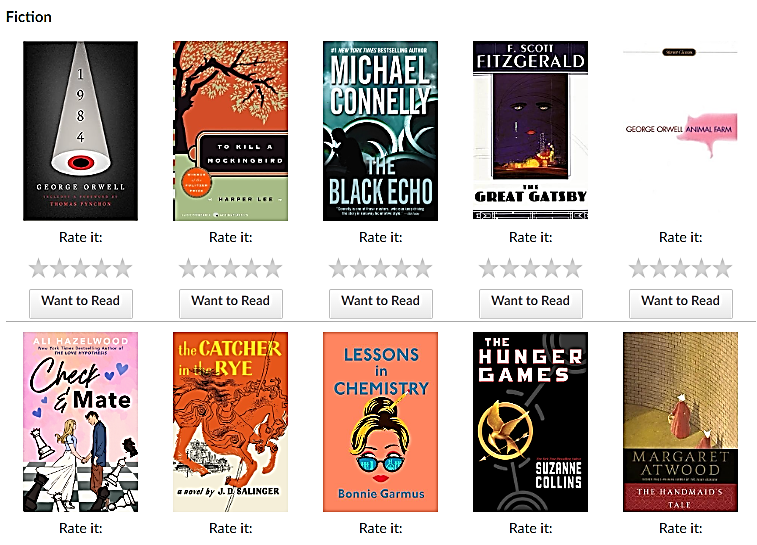


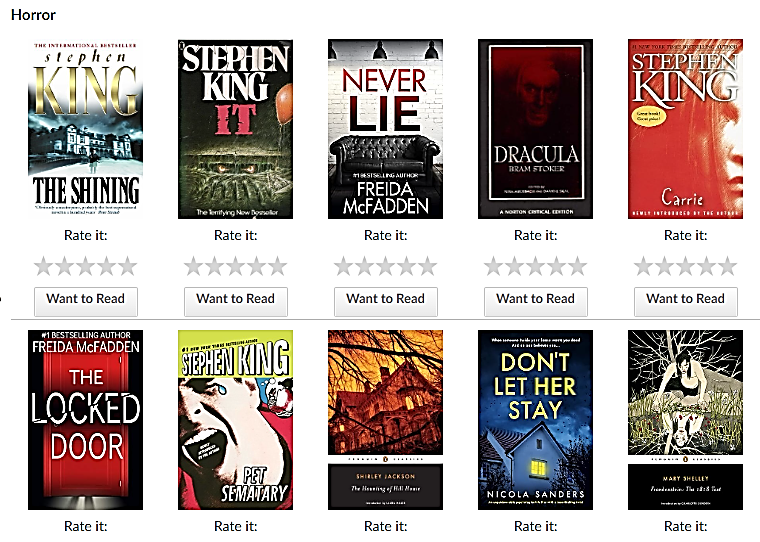
**Goodreads ask to rate at least 20 books to get started.**

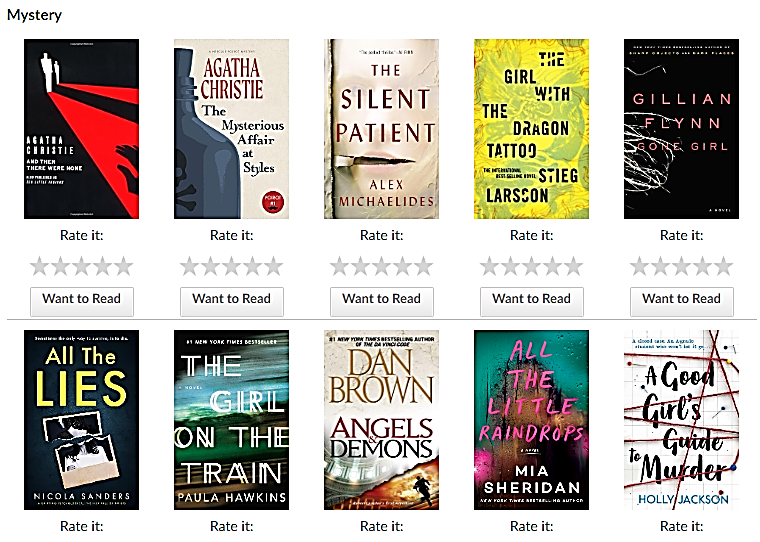


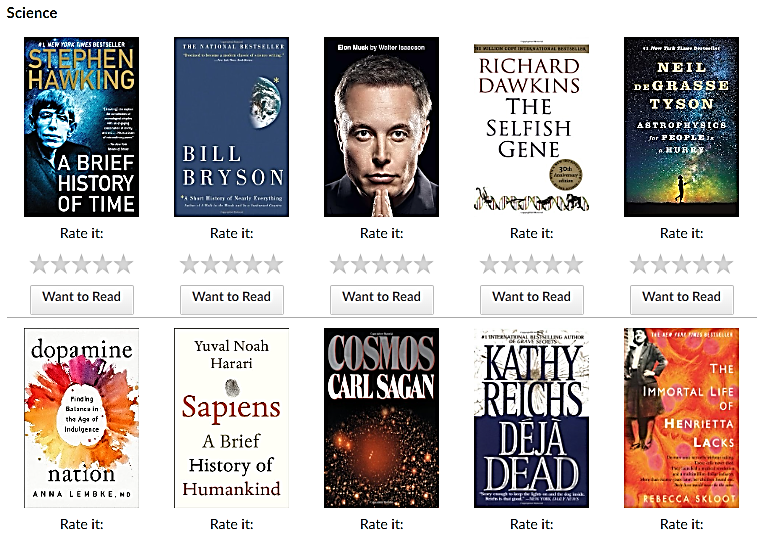
**Some of the items on specific genres offered to rate right after signing up.**

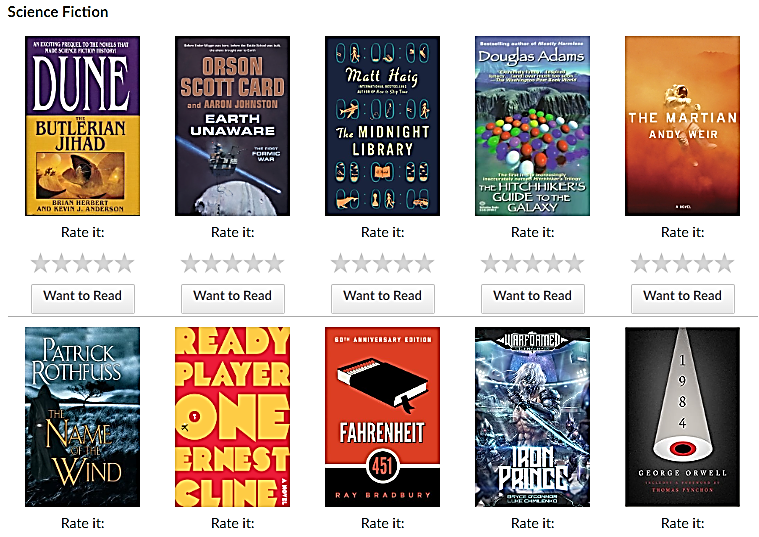




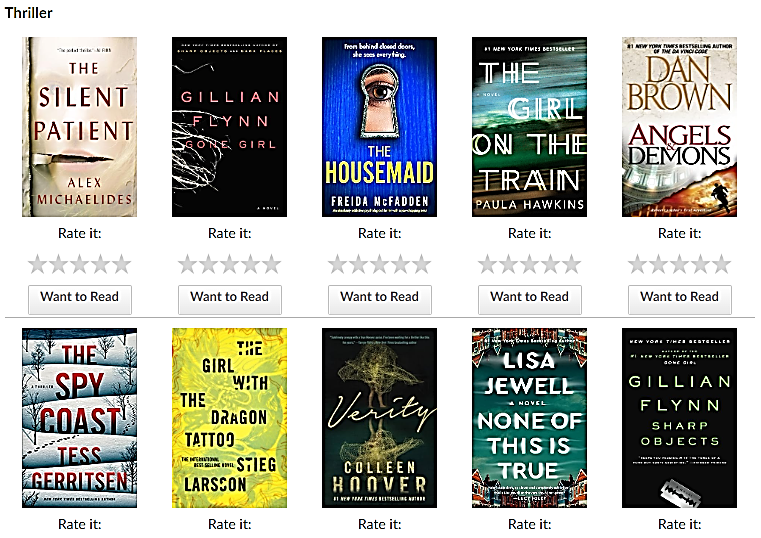


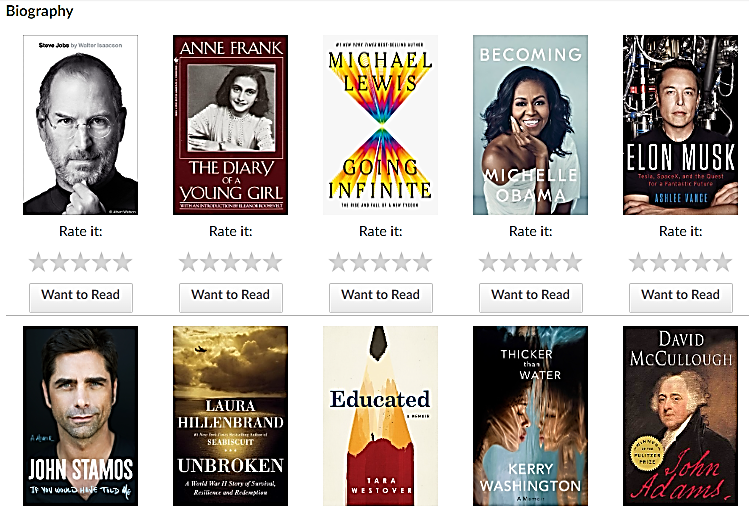












**Recommendation Engine**



Case 1: User not logged in – Generic Recommendation

Top 20 books of all time - Done

**About Dataset**

**Website:** UCSD Book Graph

* Link1: <https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home>
* Link2: <https://mengtingwan.github.io/data/goodreads.html>

**Associated GitHub Account:** <https://github.com/MengtingWan/goodreads>

**Overview:**

These datasets were collected in late 2017 from goodreads.com, where we only scraped users' public shelves, i.e. everyone can see it on web without login. User IDs and review IDs are anonymized. We collected these datasets for academic use only. Please do not redistribute them or use for commercial purposes.

We collected three groups of datasets: (1) meta-data of the books, (2) user-book interactions (users' public shelves) and (3) users' detailed book reviews. These datasets can be merged together by joining on book/user/review ids.

**Basic Statistics of the Complete Book Graph:**

* 2,360,655 books (1,521,962 works, 400,390 book series, 829,529 authors)
* 876,145 users; 228,648,342 user-book interactions in users' shelves (include 112,131,203 reads and 104,551,549 ratings)

**Meta-Data of Books**

* Detailed book graph (~2gb, about 2.3m books): goodreads\_books.json.gz
* Detailed information of authors: goodreads\_book\_authors.json.gz
* Detailed information of works (i.e., the abstract version of a book regardless any particular editions): goodreads\_book\_works.json.gz
* Detailed information of book series (Note: Unfortunately, the series id included here cannot be used for URL hack): goodreads\_book\_series.json.gz
* Extracted fuzzy book genres (genre tags are extracted from users' popular shelves by a simple keyword matching process): goodreads\_book\_genres\_initial.json.gz

**Book Shelves**

* Complete user-book interactions in 'csv' format (~4.1gb): goodreads\_interactions.csv.

**Citation:**

If you are using our datasets, please kindly cite the following papers:

* Mengting Wan, Julian McAuley, "Item Recommendation on Monotonic Behavior Chains", in RecSys'18. [bibtex]
* Mengting Wan, Rishabh Misra, Ndapa Nakashole, Julian McAuley, "Fine-Grained Spoiler Detection from Large-Scale Review Corpora", in ACL'19. [bibtex]

**Contact Email:** mengting.wan at microsoft.com

**Detailed information about datasets**

**goodreads\_books.json:**

This dataset contains all the book details.

{'isbn': '0312853122',

'text\_reviews\_count': '1',

'series': [],

'country\_code': 'US',

'language\_code': '',

'popular\_shelves': [{'count': '3', 'name': 'to-read'},

{'count': '1', 'name': 'p'},

{'count': '1', 'name': 'collection'},

{'count': '1', 'name': 'w-c-fields'},

{'count': '1', 'name': 'biography'}],

'asin': '',

'is\_ebook': 'false',

'average\_rating': '4.00',

'kindle\_asin': '',

'similar\_books': [],

'description': '',

'format': 'Paperback',

'link': '<https://www.goodreads.com/book/show/5333265-w-c-fields>',

'authors': [{'author\_id': '604031', 'role': ''}],

'publisher': "St. Martin's Press",

'num\_pages': '256',

'publication\_day': '1',

'isbn13': '9780312853129',

'publication\_month': '9',

'edition\_information': '',

'publication\_year': '1984',

'url': '<https://www.goodreads.com/book/show/5333265-w-c-fields>',

'image\_url': '<https://images.gr-assets.com/books/1310220028m/5333265.jpg>',

'book\_id': '5333265',

'ratings\_count': '3',

'work\_id': '5400751',

'title': 'W.C. Fields: A Life on Film',

'title\_without\_series': 'W.C. Fields: A Life on Film'}

We have selected only those books which has got at least 16 ratings count. With that criteria we have 1,308,957 books.

In the initial processing we have normalized the book title by removing any non-alphanumeric characters. At this stage any book which has a modified title length = 0, we have removed these books. After this processing we are left with 1,302,659 books.

Then we have removed the books which has title in some other language. Modified title = 1. We have 37931 many books of this type. After this processing we are left with 1,264,728 books.