

We intend to get our data from GoodReads.com. GoodReads lists many many genres and the genre we have selected is called “heroic fantasy.” The genre of heroic fantasy can be found here: <http://www.goodreads.com/genres/heroic-fantasy>. This genre has 2176 titles on Good Reads. According to Good Reads, Heroic fantasy is a sub-genre of fantasy which chronicles the tales of heroes in imaginary lands.

It categorizes this sub-genre by saying that the protagonist is frequently reluctant to be a champion, and/or is of low or humble origin, may have royal ancestors or parents but does not know it. Though events are usually beyond their control, the characters are thrust into positions of great responsibility where their mettle is tested in a number of spiritual and physical challenges.

We intend to write a Python script to extract the necessary information from a GoodReads review. Each title’s page has tons of information: title, author, GoodReads average rating, number of GoodReads ratings, number of GoodReads reviews, a plot summary, the number of pages in the book, genres the book subscribes to, etc. We intend to scrape sufficient information, not limited to title, author, and plot summary.