# Gutenberg

Code to read and process the Project Gutenberg catalog

The purpose of this repository is to read the Project Gutenberg catalog and parse it into a usable form. At a minimum, the intent is to create a dataframe of books, consisting of: \* Title \* Author \* Publication Date (or year) \* Language \* Subject \* Release Date \* Gutenberg book number \* Gutenberg Link

Other useful information \* Size of book file \* SF flag \* Audio flag \* Librivox link (if available) \* Other information from Librivox

My goal in this is to help me find suitable texts to read for Librivox, hence the links with Librivox (to identify what hasn’t yet been recorded), length (to find shorter books) and the SF flag (my personal interest)