# Report(part)

## 1.Data collection and clean

First, we begin to collect data, we write a spider by python to scrip data from <http://www.gutenberg.org/ebooks/search/?sort_order=release_date> and clean data.