In order to gather data on appropriate books and movies, a list of books and their film adaptations was needed. This list was acquired via web-scraping from the following URLs on Wikipedia:

<https://en.wikipedia.org/wiki/List_of_fiction_works_made_into_feature_films_%280-9%29_%26_%28A-C%29>

<https://en.wikipedia.org/wiki/List_of_fiction_works_made_into_feature_films_%28D-J%29>

<https://en.wikipedia.org/wiki/List_of_fiction_works_made_into_feature_films_%28K-R%29>

<https://en.wikipedia.org/wiki/List_of_fiction_works_made_into_feature_films_%28S-Z%29>

Java web-scrapers, see the ScraperMain, WikiScraper, and BasicScraper source files attached code documents, were used to pull the html source lines from these web pages. These html source lines were then scraped by pattern matching to identify table rows, table columns, and the book and movie titles within those table columns. Certain character patterns were identified before each book title and each movie title. This allowed those titles to be extracted and associated with each other via the table row identification. These book and movie titles were then written to a csv file and manually proofread for erroneous or ambiguous data. This index file of books and their film adaptations provided a reference point that enable targetted web scraping for relevant book and movie data.