# Web Scraping_Selenium

November 20, 2020

```
[ ]: !pip3 install selenium
     !pip install chromedriver-binary
     !pip3 install spacy
     !python -m spacy download en_core_web_sm
```

```
[1]: from bs4 import BeautifulSoup as bs
     import requests
     from selenium import webdriver
     import chromedriver_binary
     import spacy
```

```
[2]: driver = webdriver.Chrome()
     driver.page_source
     nlp = spacy.load('en_core_web_sm')
```

```
[3]: def book_scraping(href, driver, nlp):
         driver.get(href)
         soup = bs(driver.page_source, features='lxml')
         booktitle = soup.find_all('h1', id='bookTitle')[0].contents[0].strip()
         #series = soup.find_all('h2', id='bookSeries')[0].contents[1].contents[0].
     ↪strip()
         author = soup.find_all('span', itemprop='name')[0].contents[0].strip()
         descr = soup.find_all('div', id='description')[0].contents
         descr_fil= list(filter(lambda s: s!='\n', descr))
         if len(descr_fil) == 1:
             plot = ''.join(descr_fil[0].contents[0])
         else:
             descr_fil = descr_fil[1:-1]
             x = [j for i in descr_fil for j in i.contents if (isinstance(j,␣
     ↪str)==True)]
             plot = ''.join(x)
         avg_rating = soup.find_all('span', itemprop='ratingValue')[0].contents[0].
     ↪strip()
         ratings_reviews = soup.find_all('a', href='#other_reviews')
         for i in ratings_reviews:
             if i.find_all('meta',itemprop='ratingCount'):
                 ratings = i.contents[2].split()[0]
```

```
        if i.find_all('meta',itemprop='reviewCount'):
            reviews = i.contents[2].split()[0]
    doc = nlp(plot)
    token_list = [token for token in doc]
    sentiment_analysis = [token for token in token_list if not token.is_stop␣
 ↪and not token.is_punct]
    return (booktitle, author, plot, avg_rating, ratings, reviews,␣
 ↪sentiment_analysis)
```

[4]: 
```
href = 'https://www.goodreads.com/book/show/9222475-infernal-devices'
```

[5]: 
```
book_scraping(href, driver, nlp)
```

[5]: 
```
('Infernal Devices',
 'K.W. Jeter',
 "HE INHERITED A WATCHMAKER'S STORE - AND A WHOLE HEAP OF TROUBLE. But idle
sometime-musician George has little talent for clockwork. And when a shadowy
figure tries to steal an old device from the premises, George finds himself
embroiled in a mystery of time travel, music and sexual intrigue. A genuine lost
classic, a steampunk original whose time has come.",
 '3.36',
 '2,698',
 '387',
 [INHERITED,
  WATCHMAKER,
  STORE,
  HEAP,
  TROUBLE,
  idle,
  musician,
  George,
  little,
  talent,
  clockwork,
  shadowy,
  figure,
  tries,
  steal,
  old,
  device,
  premises,
  George,
  finds,
  embroiled,
  mystery,
  time,
  travel,
```

```
    music,
    sexual,
    intrigue,
    genuine,
    lost,
    classic,
    steampunk,
    original,
    time,
    come])
```

[6]: `href = 'https://www.goodreads.com/book/show/1137215.Boneshaker'`

[7]: `book_scraping(href, driver, nlp)`

[7]: ('Boneshaker',
 'Cherie Priest',
 'In the early days of the Civil War, rumors of gold in the frozen Klondike
 brought hordes of newcomers to the Pacific Northwest. Anxious to compete,
 Russian prospectors commissioned inventor Leviticus Blue to create a great
 machine that could mine through Alaska's ice. Thus was Dr. Blue's Incredible
 Bone-Shaking Drill Engine born.But on its first test run the Boneshaker went
 terribly awry, destroying several blocks of downtown Seattle and unearthing a
 subterranean vein of blight gas that turned anyone who breathed it into the
 living dead.Now it is sixteen years later, and a wall has been built to enclose
 the devastated and toxic city. Just beyond it lives Blue's widow, Briar Wilkes.
 Life is hard with a ruined reputation and a teenaged boy to support, but she and
 Ezekiel are managing. Until Ezekiel undertakes a secret crusade to rewrite
 history.His quest will take him under the wall and into a city teeming with
 ravenous undead, air pirates, criminal overlords, and heavily armed refugees.
 And only Briar can bring him out alive.',
 '3.51',
 '31,406',
 '4,073',
 [early,
  days,
  Civil,
  War,
  rumors,
  gold,
  frozen,
  Klondike,
  brought,
  hordes,
  newcomers,
  Pacific,
  Northwest,
```

Anxious,
compete,
Russian,
prospectors,
commissioned,
inventor,
Leviticus,
Blue,
create,
great,
machine,
Alaska,
ice,
Dr.,
Blue,
Incredible,
Bone,
Shaking,
Drill,
Engine,
born,
test,
run,
Boneshaker,
went,
terribly,
awry,
destroying,
blocks,
downtown,
Seattle,
unearthing,
subterranean,
vein,
blight,
gas,
turned,
breathed,
living,
dead,
sixteen,
years,
later,
wall,
built,
enclose,
devastated,

```
          toxic,
          city,
          lives,
          Blue,
          widow,
          Briar,
          Wilkes,
          Life,
          hard,
          ruined,
          reputation,
          teenaged,
          boy,
          support,
          Ezekiel,
          managing,
          Ezekiel,
          undertakes,
          secret,
          crusade,
          rewrite,
          history,
          quest,
          wall,
          city,
          teeming,
          ravenous,
          undead,
          air,
          pirates,
          criminal,
          overlords,
          heavily,
          armed,
          refugees,
          Briar,
          bring,
          alive])
```

[ ]: