Lab Exercise 3

Bansara

2024-02-28

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```r
library(rvest)
library(polite)
library(httr)
library(selectr)
```

10 movie reviews

```r
moviereviews = data.frame()

url <- "https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv"

 session <- bow(url,
                user_agent = "For Educational Purpose")


  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

 movie_title <- rep("The Shawshank Redemption",10)

  reviewer <- scrapeNodes("span.display-name-link")
  reviewer <- reviewer[1:10]

  rating_title <- scrapeNodes("a.title")
  rating_title <- rating_title[1:10]

  movie_reviews <- scrapeNodes("div.content")
  movie_reviews <- movie_reviews[1:10]
```

```r
review_date <- scrapeNodes("span.review-date")
review_date <- review_date[1:10]

review_ratings <- scrapeNodes("div.ipl-ratings-bar")
review_ratings <- review_ratings[1:10]




moviereviews <- rbind(moviereviews, data.frame(movie = movie_title,
                                    name = reviewer,
                                    title = rating_title,
                                    reviews = movie_reviews,
                                    date = review_date,
                                    ratings = review_ratings))

write.csv(moviereviews, file = "moviereviews.csv")
```