

Lab_Exercise#3_Echaveria

Luigi Echaveria

2024-03-16

```
#install.packages("dplyr")
#install.packages("rvest")
#install.packages("polite")
#install.packages("httr")

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

library(rvest)
library(polite)
library(httr)

reviews <- data.frame()

link <- "https://www.imdb.com/title/tt0068646/reviews?ref_=tt_urv"
session <- bow(link, user_agent = "Educational Purpose")

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

titleMovie <- scrapeNodes("a.subnav_heading")
reviewer <- scrapeNodes("span.display-name-link")[1:10]
scrapedDate <- scrapeNodes("span.review-date")[1:10]
scrapedRating <- scrapeNodes("span.rating-other-user-rating")[1:10]
scrapedReview <- scrapeNodes("div.text.show-more__control")[1:10]
scrapedTitleReview <- scrapeNodes("a.title")[1:10]

reviews <- rbind(reviews, data.frame(
  movie_title = rep(titleMovie, 10),
  reviewer_name = reviewer,
```

```
review_date = scrapedDate,  
review_rating = scrapedRating,  
review_content = scrapedReview,  
review_title = scrapedTitleReview  
)
```

reviews

```
##      movie_title      reviewer_name      review_date review_rating  
## 1  The Godfather      CalRhys      20 June 2014      10/10  
## 2  The Godfather andrewburgereviews      1 April 2019      10/10  
## 3  The Godfather      Sleepin_Dragon  6 February 2021      10/10  
## 4  The Godfather      gogoschka-1 11 February 2018      10/10  
## 5  The Godfather      OU 23 February 2020      10/10  
## 6  The Godfather      alexkolokotronis      21 June 2008      10/10  
## 7  The Godfather      b-a-h TNT-6      5 March 2002      10/10  
## 8  The Godfather      planktonrules      30 March 2013      10/10  
## 9  The Godfather      TheLittleSongbird      18 July 2010      10/10  
## 10 The Godfather      mattrochman 14 October 2006      10/10  
##  
## 1  
## 2  
## 3  
## 4  
## 5  
## 6
```