RWorksheet#5

2023-12-01

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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(polite)
library(xml2)
library(magrittr)
library(rvest)
library(httr)


#Movie Guide
#m1 - Breaking Bad
#m2 - Game of Thrones
#m3 - Arcane
#m4 - Death Note
#m5 - Better Call Saul

polite::use_manners(save_as = "polite_scrape.R")
```

```
## v Setting active project to 'F:/RScraping'
```

```
url_m1 <- 'https://www.imdb.com/title/tt0903747/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilter=
url_m2 <- 'https://www.imdb.com/title/tt0944947/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilter=
url_m3 <- 'https://www.imdb.com/title/tt11126994/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilte
```

```
url_m4 <- 'https://www.imdb.com/title/tt0877057/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilter=
url_m5 <- 'https://www.imdb.com/title/tt3032476/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilter=
```

```
session_m1 <- bow(url_m1,
                  user_agent = "Educational")
session_m2 <- bow(url_m2,
                  user_agent = "Educational")
session_m3 <- bow(url_m3,
                  user_agent = "Educational")
session_m4 <- bow(url_m4,
                  user_agent = "Educational")
session_m5 <- bow(url_m5,
                  user_agent = "Educational")

session_m1
```

```
## <polite session> https://www.imdb.com/title/tt0903747/reviews?spoiler=hide&sort=curated&dir=desc&rati
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
session_m2
```

```
## <polite session> https://www.imdb.com/title/tt0944947/reviews?spoiler=hide&sort=curated&dir=desc&rati
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
session_m3
```

```
## <polite session> https://www.imdb.com/title/tt11126994/reviews?spoiler=hide&sort=curated&dir=desc&ra
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
session_m4
```

```
## <polite session> https://www.imdb.com/title/tt0877057/reviews?spoiler=hide&sort=curated&dir=desc&rati
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
session_m5
```

```
## <polite session> https://www.imdb.com/title/tt3032476/reviews?spoiler=hide&sort=curated&dir=desc&rati
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
reviewerName_m1 <- character(0)
dateReviewed_m1 <- character(0)
userRating_m1 <- character(0)
titleReview_m1 <- character(0)
textReview_m1 <- character(0)

reviewerName_m2 <- character(0)
dateReviewed_m2 <- character(0)
userRating_m2 <- character(0)
titleReview_m2 <- character(0)
textReview_m2 <- character(0)

reviewerName_m3 <- character(0)
dateReviewed_m3 <- character(0)
userRating_m3 <- character(0)
titleReview_m3 <- character(0)
textReview_m3 <- character(0)

reviewerName_m4 <- character(0)
dateReviewed_m4 <- character(0)
userRating_m4 <- character(0)
titleReview_m4 <- character(0)
textReview_m4 <- character(0)

reviewerName_m5 <- character(0)
dateReviewed_m5 <- character(0)
userRating_m5 <- character(0)
titleReview_m5 <- character(0)
textReview_m5 <- character(0)


#Breaking Bad
tv_m1 <- scrape(session_m1) %>%
  html_elements('div.lister-item')

reviewerName_m1 <- tv_m1 %>%
  html_nodes('span.display-name-link') %>%
  html_text()

dateReviewed_m1 <- tv_m1 %>%
  html_nodes('span.review-date') %>%
  html_text()

userRating_m1 <- tv_m1 %>%
  html_node(".rating-other-user-rating") %>%
  html_text()

titleReview_m1 <- tv_m1 %>%
```

```
  html_nodes('a.title') %>%
  html_text()

textReview_m1 <- tv_m1 %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()

DF_m1 <- data.frame(userRating_m1, dateReviewed_m1, reviewerName_m1, titleReview_m1, textReview_m1)
colnames(DF_m1) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review")


#Game of Thrones

tv_m2 <- scrape(session_m2) %>%
  html_elements('div.lister-item')

reviewerName_m2 <- tv_m2 %>%
  html_nodes('span.display-name-link') %>%
  html_text()

dateReviewed_m2 <- tv_m2 %>%
  html_nodes('span.review-date') %>%
  html_text()

userRating_m2 <- tv_m2 %>%
  html_node(".rating-other-user-rating") %>%
  html_text()

titleReview_m2 <- tv_m2 %>%
  html_nodes('a.title') %>%
  html_text()

textReview_m2 <- tv_m2 %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()


DF_m2 <- data.frame(userRating_m2, dateReviewed_m2, reviewerName_m2, titleReview_m2, textReview_m2)
colnames(DF_m2) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review")
View(DF_m2)


#Arcane
tv_m3 <- scrape(session_m3) %>%
  html_elements('div.lister-item')

reviewerName_m3 <- tv_m3 %>%
  html_nodes('span.display-name-link') %>%
  html_text()

dateReviewed_m3 <- tv_m3 %>%
  html_nodes('span.review-date') %>%
  html_text()
```

```r
userRating_m3 <- tv_m3 %>%
  html_node(".rating-other-user-rating") %>%
  html_text()

titleReview_m3 <- tv_m3 %>%
  html_nodes('a.title') %>%
  html_text()

textReview_m3 <- tv_m3 %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()

DF_m3 <- data.frame(userRating_m3, dateReviewed_m3, reviewerName_m3, titleReview_m3, textReview_m3)
colnames(DF_m3) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review")

View(DF_m3)

#Death Note

tv_m4 <- scrape(session_m4) %>%
  html_elements('div.lister-item')

reviewerName_m4 <- tv_m4 %>%
  html_nodes('span.display-name-link') %>%
  html_text()

dateReviewed_m4 <- tv_m4 %>%
  html_nodes('span.review-date') %>%
  html_text()

userRating_m4 <- tv_m4 %>%
  html_node(".rating-other-user-rating") %>%
  html_text()

titleReview_m4 <- tv_m4 %>%
  html_nodes('a.title') %>%
  html_text()

textReview_m4 <- tv_m4 %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()

DF_m4 <- data.frame(userRating_m4, dateReviewed_m4, reviewerName_m4, titleReview_m4, textReview_m4)
colnames(DF_m4) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review")

View(DF_m4)

#Better Call Saul

tv_m5 <- scrape(session_m5) %>%
  html_elements('div.lister-item')

reviewerName_m5 <- tv_m5 %>%
```

```
  html_nodes('span.display-name-link') %>%
  html_text()

dateReviewed_m5 <- tv_m5 %>%
  html_nodes('span.review-date') %>%
  html_text()

userRating_m5 <- tv_m5 %>%
  html_node(".rating-other-user-rating") %>%
  html_text()

titleReview_m5 <- tv_m5 %>%
  html_nodes('a.title') %>%
  html_text()

textReview_m5 <- tv_m5 %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()

DF_m5 <- data.frame(userRating_m5, dateReviewed_m5, reviewerName_m5, titleReview_m5, textReview_m5)
colnames(DF_m5) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review")

View(DF_m5)
```