part2amazonlabact

Quillo

2024-02-08

libraries

```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)
library(xml2)
```

First product(table):

```r
reviewsofLight = data.frame()


link01 <- "https://www.amazon.com.au/s?k=table&page=2&qid=1707956110&ref=sr_pg_2"

link02 <- "https://www.amazon.com.au/s?k=table&page=3&qid=1707956125&ref=sr_pg_3"

link03 <- "https://www.amazon.com.au/s?k=table&page=4&qid=1707956155&ref=sr_pg_4"

link04 <- "https://www.amazon.com.au/s?k=table&page=5&qid=1707956166&ref=sr_pg_5"

link05 <- "https://www.amazon.com.au/s?k=table&page=6&qid=1707956172&ref=sr_pg_6"

link06 <- "https://www.amazon.com.au/s?k=table&page=7&qid=1707956186&ref=sr_pg_7"

  session <- bow(link01,
                 user_agent = "For Educational Purpose")

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

```r
  }

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]



  products1 <- rbind(reviewsofLight, data.frame(
                    name = product_name,
                    price = product_price,
                    ratings = product_ratings,
                    reviews = product_reviews))

  Sys.sleep(5)

#link 2

session <- bow(link02,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]



  products2 <- rbind(reviewsofLight, data.frame(
                    name = product_name,
                    price = product_price,
                    ratings = product_ratings,
                    reviews = product_reviews))

#link 3

session <- bow(link03,
               user_agent = "For Educational Purpose")
```

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

product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
product_name <- product_name[1:45]
product_price <- scrapeNodes("span.a-offscreen")
product_price <- product_price[1:45]
product_ratings <- scrapeNodes("span.a-icon-alt")
product_ratings <- product_ratings[1:45]
product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
product_reviews <- product_reviews[1:45]




products3 <- rbind(reviewsofLight, data.frame(
                   name = product_name,
                   price = product_price,
                   ratings = product_ratings,
                   reviews = product_reviews))

Sys.sleep(5)

#link 4

session <- bow(link03,
               user_agent = "For Educational Purpose")

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

product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
product_name <- product_name[1:45]
product_price <- scrapeNodes("span.a-offscreen")
product_price <- product_price[1:45]
product_ratings <- scrapeNodes("span.a-icon-alt")
product_ratings <- product_ratings[1:45]
product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
product_reviews <- product_reviews[1:45]




products4 <- rbind(reviewsofLight, data.frame(
                   name = product_name,
                   price = product_price,
                   ratings = product_ratings,
                   reviews = product_reviews))
```

```r
#link 5

session <- bow(link05,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]




  products5 <- rbind(reviewsofLight, data.frame(
                     name = product_name,
                     price = product_price,
                     ratings = product_ratings,
                     reviews = product_reviews))

    Sys.sleep(5)
#link 6

session <- bow(link06,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]



  products6<- rbind(reviewsofLight, data.frame(
                     name = product_name,
```

```
                        price = product_price,
                        ratings = product_ratings,
                        reviews = product_reviews))



finalproduct = rbind(products1, products2, products3, products4, products5, products6)

 Sys.sleep(5)
```

Second product(chair):

```
library(dplyr)
library(rvest)
library(polite)
library(httr)
library(selectr)
library(xml2)

reviewsofLight = data.frame()


link01 <- "https://www.amazon.com.au/s?k=chair&page=2&crid=2P3P0Y4SK8EGW&qid=1707956287&sprefix=chai%2C

link02 <- "https://www.amazon.com.au/s?k=chair&page=3&crid=2P3P0Y4SK8EGW&qid=1707956291&sprefix=chai%2C

link03 <- "https://www.amazon.com.au/s?k=chair&page=4&crid=2P3P0Y4SK8EGW&qid=1707956304&sprefix=chai%2C

link04 <- "https://www.amazon.com.au/s?k=chair&page=5&crid=2P3P0Y4SK8EGW&qid=1707956311&sprefix=chai%2C

link05 <- "https://www.amazon.com.au/s?k=chair&page=6&crid=2P3P0Y4SK8EGW&qid=1707956318&sprefix=chai%2C

link06 <- "https://www.amazon.com.au/s?k=chair&page=7&crid=2P3P0Y4SK8EGW&qid=1707956328&sprefix=chai%2C

  session <- bow(link01,
                 user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]
```

```r
  products1 <- rbind(reviewsofLight, data.frame(
                    name = product_name,
                    price = product_price,
                    ratings = product_ratings,
                    reviews = product_reviews))
 Sys.sleep(5)
#link 2

session <- bow(link02,
              user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]




  products2 <- rbind(reviewsofLight, data.frame(
                    name = product_name,
                    price = product_price,
                    ratings = product_ratings,
                    reviews = product_reviews))
 Sys.sleep(5)
#link 3

session <- bow(link03,
              user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]
```

```r
  products3 <- rbind(reviewsofLight, data.frame(
                     name = product_name,
                     price = product_price,
                     ratings = product_ratings,
                     reviews = product_reviews))
 Sys.sleep(5)
#link 4

session <- bow(link03,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]




  products4 <- rbind(reviewsofLight, data.frame(
                     name = product_name,
                     price = product_price,
                     ratings = product_ratings,
                     reviews = product_reviews))
 Sys.sleep(5)
#link 5

session <- bow(link05,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
```

```r
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]



  products5 <- rbind(reviewsofLight, data.frame(
                     name = product_name,
                     price = product_price,
                     ratings = product_ratings,
                     reviews = product_reviews))
 Sys.sleep(5)
#link 6

session <- bow(link06,
               user_agent = "For Educational Purpose")

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

  product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
  product_name <- product_name[1:45]
  product_price <- scrapeNodes("span.a-offscreen")
  product_price <- product_price[1:45]
  product_ratings <- scrapeNodes("span.a-icon-alt")
  product_ratings <- product_ratings[1:45]
  product_reviews <- scrapeNodes("span.a-size-base.s-underline-text")
  product_reviews <- product_reviews[1:45]



  products6<- rbind(reviewsofLight, data.frame(
                     name = product_name,
                     price = product_price,
                     ratings = product_ratings,
                     reviews = product_reviews))



finalproduct2 = rbind(products1, products2, products3, products4, products5, products6)
```