Using `RSelenium` for web scraping

Example 5. Product price scraping from https://www.daraz.com.np/smartphones

---

In [ ]:
```
library(rvest) #see https://rvest.tidyverse.org/articles/harvesting-the-web.html for details
library(dplyr)
library(RSelenium)
library(netstat)
library(httr)
```

In [2]:
```
rD <- rsDriver(browser = "firefox", port = free_port())
remDr <- rD$client
remDr$navigate("https://www.daraz.com.np/smartphones")
```

In [3]:
```
product <- c()
rating <- c()
rating_no <- c()
sales <- c()
price <- c()

web_scrap <- function() {
    webpage <- read_html(remDr$getPageSource()[[1]])
    elems <- webpage %>% html_nodes(xpath = "//div[starts-with(@class,'description')]")
        for (e in elems) {
            val <- e %>% html_nodes(xpath = "div[1]") %>% html_text(trim=TRUE)
            product <<- c(product, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[2]//span[2]") %>% html_text(trim=TRUE)
            rating <<- c(rating, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[2]//span[3]") %>% html_text(trim=TRUE)
            rating_no <<- c(rating_no, ifelse(length(val) == 0,"",val))
```

```
            val <- e %>% html_nodes(xpath = "div[2]/div[3]") %>% html_text(trim=TRUE)
            sales <<- c(sales, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[@id='id-price']//div[starts-with(@class,'current-price')]") %>% html_text(tri
            price <<- c(price, ifelse(length(val) == 0,"",val))
        }
    }
```

In [4]:
```
#navigating pages from 1 to 3 and scraping data
for (i in 1:3) {
    elem <- remDr$findElement(using = "xpath", value = sprintf('//li[@title = "%s" ]', i))
    elem$clickElement()
    web_scrap()
    }
```

In [5]:
```
df <- data.frame(product, rating, rating_no, sales, price)
head(df)
write.csv(df,file = 'example5.csv')
```

A data.frame: 6 × 5

	product	rating	rating_no	sales	price
	<chr>	<chr>	<chr>	<chr>	<chr>
1	Tecno Spark 20 Pro+ (16*/256 GB) | 6.78" FHD + AMOLED Curved Screen | 120Hz Refresh Rate | 100 Days Replacement Warranty | G99 Ultra Boost Processor | 108MP Ultra Sensing Main Camera | 5000mAh Battery | 33W Super Charge	4.5/5	(39)	127 Sold	Rs.26,990
2	Redmi Note 11 | 90 Hz FHD+ AMOLED Display | 50 MP AI Quad Camera | 4/64 GB - Black	4.5/5	(115)	377 Sold	Rs.23,999
3	Redmi 13C (6/128GB) | 6.74" Dot Drop display | 90Hz Refresh Rate | 5000mAh Battery | 18W PD charging	4.4/5	(36)	156 Sold	Rs.15,999
4	realme C53 (6+128 GB) | 6.74 inch HD+ IPS LCD Display | 108MP Back Camera | 5000mAh Battery with 18W Quick charge	4.1/5	(18)	85 Sold	Rs.16,499
5	Redmi Note 13 Pro (8/256GB) | 6.67" AMOLED Display | MediaTek Helio G99-Ultra Processor | 5000mAh Battery | 67W Turbo Charging	4.3/5	(19)	80 Sold	Rs.32,999
6	realme C51 (4+64 GB) | 5000mAh with 33W SUPERVOOC charge | Back Camera(s): 50MP (Samsung HM6) | 90Hz Refresh Rate |	4.4/5	(8)	40 Sold	Rs.13,499

Practice 3. From https://www.sharesansar.com/today-share-price, scrape stock data of Commercial Bank from date 2024-06-06 to 2024-06-11

In [6]:
```
#loading the website
remDr$navigate("https://www.sharesansar.com/today-share-price")
```

In [7]:
```
#clicking on the dropdown box of sector
elem <- remDr$findElement(using = "xpath", "//span[@id='select2-sector-container']")
elem$clickElement()

#Finding input field to type
elem <- remDr$findElement(using = "xpath", "//input[@role='textbox' and @type='search']")
elem$sendKeysToElement(list("Commercial Bank")) #typing Commercial Bank
```

```
elem$sendKeysToElement(list(key = "enter")) #sending Enter key signal

#list of date to scrape
lst_date <- c('2024-06-06','2024-06-07','2024-06-08','2024-06-09', '2024-06-10', '2024-06-11')

#removing any previous dataframe named df_stock
if (exists('df_stock')){
    rm(df_stock)
}

for (l in lst_date) {
    elem <- remDr$findElement(using = "xpath", "//input[@name='date']")
    elem$clearElement() #clearing the input filed
    elem$sendKeysToElement(list(l)) #entering the date
    elem$sendKeysToElement(list(key = "enter")) #sending Enter key signal

    elem <- remDr$findElement(using = "xpath", '//button[@id="btn_todayshareprice_submit"]')
    elem$clickElement()
    Sys.sleep(3) #waiting 3 seconds to allow the page to fully load

    webpage <- read_html(remDr$getPageSource()[[1]]) #obtaining html code from the page

    if (grepl('No Record Found.', webpage) == FALSE) { #checking whether No record found is displayed in the page or not
        tbl <- webpage %>% html_table()
        dd <- tbl[[2]] #the required data table is in the second index
        dd$date_en <- l #adding a date column

        if (exists('df_stock')){
            df_stock <- rbind(df_stock, dd)
        } else {
            df_stock <- dd
        }
    }
}
```

In [8]:
```
head(df_stock)
write.csv(df_stock, file="practice3.csv", row.names=F)
```

A tibble: 6 × 20

S.No	Symbol	Conf.	Open	High	Low	Close	VWAP	Vol	Prev. Close	Turnover	Trans.	Diff	Range	Diff %	Range %	VWA %
<int>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	ADBL	44.64	268.5	268.5	260.0	261.9	260.75	23,243.00	264.0	6,060,520.50	168	-2.1	8.5	-0.80	3.27	0.4
2	CZBIL	39.02	167.1	168.0	165.2	167.0	166.06	29,887.00	168.0	4,963,024.60	83	-1.0	2.8	-0.60	1.69	0.5
3	EBL	45.35	528.1	528.1	521.2	528.0	524.52	27,100.00	528.9	14,214,538.10	144	-0.9	6.9	-0.17	1.32	0.6
4	GBIME	39.92	183.0	184.0	180.1	180.5	180.89	64,245.00	183.9	11,621,120.20	340	-3.4	3.9	-1.85	2.17	-0.2
5	HBL	41.49	190.0	190.0	185.0	187.5	186.92	27,461.00	189.5	5,133,072.20	134	-2.0	5.0	-1.06	2.70	0.3
6	KBL	38.03	143.8	143.8	138.3	140.0	139.14	59,331.00	141.0	8,255,146.10	333	-1.0	5.5	-0.71	3.98	0.6

Practice 4. From https://www.daraz.com.np search for top selling `rice` products. Then, scrape rice prices from 1 to 5 pages

In [9]:
```
remDr$navigate("https://www.daraz.com.np")

product <- c()
rating <- c()
rating_no <- c()
sales <- c()
current_price <- c()
```

```
original_price <- c()

web_scrap <- function() {
    webpage <- read_html(remDr$getPageSource()[[1]])
    elems <- webpage %>% html_nodes(xpath = "//div[starts-with(@class,'description')]")
        for (e in elems) {
            val <- e %>% html_nodes(xpath = "div[1]") %>% html_text(trim=TRUE)
            product <<- c(product, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[2]//span[2]") %>% html_text(trim=TRUE)
            rating <<- c(rating, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[2]//span[3]") %>% html_text(trim=TRUE)
            rating_no <<- c(rating_no, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[2]/div[3]") %>% html_text(trim=TRUE)
            sales <<- c(sales, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[@id='id-price']//div[starts-with(@class,'current-price')]") %>% html_text(tri
            current_price <<- c(current_price, ifelse(length(val) == 0,"",val))

            val <- e %>% html_nodes(xpath = "div[@id='id-price']//div[starts-with(@class,'original-price')]") %>% html_text(tr
            original_price <<- c(original_price, ifelse(length(val) == 0,"",val))
        }
    }


#searching for rice products
elem <- remDr$findElement(using = "xpath", "//input[@id='q']")
elem$clearElement() #clearing the input filed
elem$sendKeysToElement(list('rice')) #entering the date
elem$sendKeysToElement(list(key = "enter")) #sending Enter key signal
Sys.sleep(5) #waiting 5 second for

#sort by Top Sales
remDr$findElement(using = "xpath", "//div[@role='combobox']")$clickElement()
remDr$findElement(using = "xpath", "//li[@title='Top Sales']")$clickElement()
```

In [10]:
```
#navigating pages from 1 to 5 and scraping data
for (i in 1:5) {
```

```
    elem <- remDr$findElement(using = "xpath", value = sprintf('//li[@title = "%s" ]', i))
    elem$clickElement()
    web_scrap()
    Sys.sleep(1)
    }
```

In [11]:
```
df <- data.frame(product, rating, rating_no, sales, current_price, original_price)
head(df)
write.csv(df,file = 'practice4.csv', row.names = F)
```

A data.frame: 6 × 6

	product	rating	rating_no	sales	current_price	original_price
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	Lal Qilla Brown Basmati Rice 1 kg	4.4/5	(7)	31 Sold	Rs.410	Rs. 500
2	DhikiJato Local Anadi Chamal 1 KG	4.5/5	(20)	145 Sold	Rs.350	
3	Dhiki Jato Jumla Marsi Chamal 1kg	4.4/5	(10)	51 Sold	Rs.270	
4	Newari Shahi Pulao Basmati Rice 5 Kg	4.9/5	(14)	94 Sold	Rs.995	
5	Taichin Chamal 1Kg	4.7/5	(12)	161 Sold	Rs.195	
6	Newari Shahi Pulao Basmati Rice - 5 Kg	5/5	(2)		Rs.995	Rs. 1,000

In [12]:
```
# Close the server
remDr$close()
rD$server$stop()
```

TRUE