Mini Project 01 - IMDB web scraping

```
library(tidyverse)
library(rvest)
```

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
print(url)
```

```
[1] "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
# read html
imdb <- read_html(url)
```

```
# movie title
titles <- imdb %>%
    html_nodes("h3.lister-item-header") %>%
    html_text2()
```

```
titles[1:10]
```

'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' ·
'4. Schindler\'s List (1993)' · '5. The Godfather Part II (1974)' · '6. 12 Angry Men (1957)' ·
'7. The Lord of the Rings: The Return of the King (2003)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·
'10. Fight Club (1999)'

```
# rating
ratings <- imdb %>%
```

```
    html_nodes("div.ratings-imdb-rating") %>%
    html_text2() %>%
    as.numeric
```

```
ratings[1:10]
```

9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8

```
num_votes <- imdb %>%
    html_nodes("p.sort-num_votes-visible") %>%
    html_text2()
```

```
df <- data.frame(
    title = titles,
    rating = ratings,
    num_vote = num_votes
)

head(df)
```

A data.frame: 6 × 3

	title	rating	num_vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,697,456 | Gross: $28.34M | Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,871,876 | Gross: $134.97M | Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,671,183 | Gross: $534.86M | Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,363,761 | Gross: $96.90M | Top 250: #6
5	5. The Godfather Part II (1974)	9.0	Votes: 1,279,429 | Gross: $57.30M | Top 250: #4
6	6. 12 Angry Men (1957)	9.0	Votes: 796,598 | Gross: $4.36M | Top 250: #5

Mini Project 02 - Specphone Phone Database

```
library(tidyverse)
library(rvest)
```

```
Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
Warning message:
“Failed to locate timezone database”
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1

✓ ggplot2 3.3.5     ✓ purrr   0.3.4
✓ tibble  3.1.5     ✓ dplyr   1.0.7
✓ tidyr   1.1.4     ✓ stringr 1.4.0
✓ readr   2.0.2     ✓ forcats 0.5.1

── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
✖ dplyr::filter()  masks stats::filter()
✖ purrr::flatten() masks jsonlite::flatten()
✖ dplyr::lag()     masks stats::lag()


Attaching package: ‘rvest’
```

```
url <- read_html("https://specphone.com/Samsung-Galaxy-S23-Ultra-5G.html")
```

```
att <- url %>%
    html_nodes("div.topic") %>%
    html_text2()

value <- url %>%
    html_nodes("div.detail") %>%
    html_text2()
```

```
data.frame(attribute = att, value = value)
```

A data.frame: 33 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	**มกราคม** 2566
วันวางจำหน่าย	**ยังไม่วางจำหน่าย**
ขนาด	163.40 x 78.10 x 8.90 **มม.**
น้ำหนัก	233 **กรัม**
วัสดุ	Glass front (Gorilla Glass Victus 2), glass back (Gorilla Glass Victus 2), aluminum frame
SIM	**รองรับ** 2 **ซิมการ์ด** (nano sim, nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A (CA), 5G
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	2100/2600/3500/4700
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A (CA), 5G
ประเภท	Dynamic AMOLED 2X
ขนาดหน้าจอ	6.80 **นิ้ว**
ความละเอียด	1440 x 3088 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Qualcomm Snapdragon 8 Gen 2 SM8550 3.2 GHz
ชิปกราฟิก	Adreno 740
หน่วยความจำ	12 GB
ความจุ	512/1024 GB
Memory Card	**ไม่รองรับ**
กล้องหลัก	**ตัวที่** 1: 200 MP, f/1.7, 23mm (wide), 1/1.3 **ตัวที่** 2: 10 MP, f/4.9, 230mm (periscope telephoto), 1/3.52 **ตัวที่** 3: 10 MP, f/2.4, 70mm (telephoto), 1/3.52 **ตัวที่** 4: 12 MP, f/2.2, 13mm, 120˚ (ultrawide), 1/2.55
ความละเอียดวีดีโอ	8K@24/30fps, 4K@30/60fps, 1080p@30/60/240fps, 720p@960fps, HDR10+, stereo sound rec., gyro-EIS
กล้องหน้า	**ตัวที่** 1: 12 MP, f/2.2, 25mm (wide), PDAF
Bluetooth	5.3, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac/6e, tri
USB	Type-C
GPS	GPS, GLONASS, BDS, GALILE
NFC	**รอบรับ**
ความจ	5,000 mAh

```
apple_url <- read_html("https://specphone.com/brand/Apple")
```

```
# links to all smartphone
links <- apple_url %>%
    html_nodes("li.mobile-brand-item a") %>%
    html_attr("href")
```

```
full_links <- paste0("https://specphone.com", links)
```

```
result <- data.frame()

for (link in full_links[1:10]) {
    app_topic <- link %>%
        read_html() %>%
        html_nodes("div.topic") %>%
        html_text2()

    app_detail <- link %>%
        read_html() %>%
        html_nodes("div.detail") %>%
        html_text2()

    tmp <- data.frame(attribute = app_topic,
                      value = app_detail)

    result <- bind_rows(result, tmp)
    print("Progress ...")
}

print(result)
```

```
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
[1] "Progress ..."
```

	attribute
1	วันเปิดตัว
2	วันวางจำหน่าย
3	ขนาด
4	น้ำหนัก
5	วัสดุ
6	SIM
7	Technology
8	2G
9	3G

```
print(head(result),3)
```

	attribute	value
1	วันเปิดตัว	พฤศจิกายน 2555
2	วันวางจำหน่าย	พฤษภาคม 2556, สินค้าจำหน่ายหมดแล้ว
3	ขนาด	200.00 x 134.70 x 7.20 มม.
4	น้ำหนัก	308 กรัม
5	วัสดุ	Aluminium, Plastic
6	SIM	รองรับ 1 ซิมการ์ด

```
# write csv
write_csv(result, "result_apphone.csv")
```