**Mini Project 01 - IMDB Wed scraping **

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

```
mini
```

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

```
print(url)
```

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

```
imdb <- read_html(url)
```

imdb

```
{html_document}
<html xmlns:og="http://ogp.me/ns#" xmlns:fb="http://www.facebook.com/2008/
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UT
[2] <body id="styleguide-v2" class="fixed">\n            <img height="1" w
```

```
#movie title #text2=remove spacial charecter. #node = choose only lthe first
titles <- imdb %>%
            html_nodes("h3.lister-item-header") %>%
            html_text2()
```

```
#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

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

```
#Build a dataset
df <- data.frame(
    title = titles ,
    rating = ratings ,
    vote = votes
)
head(df)
```

A data.frame: 6 × 3

	title	rating	vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,681,787 | Gross: $28.34M | Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,859,304 | Gross: $134.97M | Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,655,144 | Gross: $534.86M | Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,848,011 | Gross: $377.85M | Top 250: #7
5	5. The Godfather Part II (1974)	9.0	Votes: 1,272,627 | Gross: $57.30M | Top 250: #4
6	6. Schindler's List (1993)	9.0	Votes: 1,356,983 | Gross: $96.90M | Top 250: #6

**Mini Project 02 - Phone Database **

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

```
url2 <- read_html("https://specphone.com/ZTE-nubia-Red-Magic-8-Pro-.html")
```

```
sp <- url2
```

```
atts <- sp %>%
    html_nodes("div.topic") %>%
    html_text2()
values <- sp %>%
    html_nodes("div.detail") %>%
    html_text2()
```

```
data.frame(attribute = atts , value= values)
```

A data.frame: 32 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ธันวาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	164.00 x 76.40 x 8.90 มม.
น้ำหนัก	230 กรัม
วัสดุ	Glass front, glass back, aluminum frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA, LTE-A, 5G
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	2100/2600/3500/4700
ความเร็ว	HSPA, LTE-A, 5G
ประเภท	AMOLED
ขนาดหน้าจอ	6.80 นิ้ว
ความละเอียด	1116 x 2480 pixels
ระบบปฏิบัติการ	Android 13
ชิปประมวลผล	Qualcomm Snapdragon 8 Gen 2 SM8550 3.2 GHz
ชิปกราฟิก	Adreno 740

หน่วยความจำ	12 GB
ความจุ	256 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), 1/1.57 ตัวที่ 2: 8 MP, f/2.2, 120˚, 13mm (ultrawide), 1/4.0 ตัวที่ 3: 2 MP, f/2.4, (macro)
ความละเอียดวีดีโอ	8K@30fps, 4K@30/60fps, 1080p@30/60/120/240fps
กล้องหน้า	ตัวที่ 1: 16 MP, (wide), under display
Bluetooth	5.3, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac/6e, dua
USB	Type-C
GPS	GPS (L1+L5), GLONASS, BDS
NFC	รอบรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (165W)

```
url3 <- read_html("https://specphone.com/brand/Samsung")
```

```
ss_link <- url3
```

```
links <- ss_link %>%
    html_nodes("li.mobile-brand-item a") %>%
    html_attr("href")
```

links

'/Samsung-Galaxy-M13.html' · '/Samsung-Galaxy-A23.html' · '/Samsung-Galaxy-A13.html' ·
'/Samsung-Galaxy-M32-5G.html' · '/Samsung-Galaxy-A12-Nacho.html' ·
'/Samsung-Galaxy-Pocket-Neo.html' · '/Samsung-Galaxy-Young.html' ·
'/Samsung-Galaxy-J1-Mini.html' · '/Samsung-Galaxy-A01-Core-1-16GB.html' ·
'/Samsung-Galaxy-V-PLUS.html' · '/Samsung-Galaxy-Young-2.html' · '/Samsung-Galaxy-M02.html' ·
'/Samsung-Galaxy-A11.html' · '/Samsung-Galaxy-J2-Pro-2018.html' ·
'/Samsung-Galaxy-A12-2021.html' · '/Samsung-Galaxy-A21s-3-32GB.html' ·
'/Samsung-Galaxy-J5.html' · '/Samsung-Galaxy-J4.html' · '/Samsung-Galaxy-Core-2-Duos.html' ·
'/Samsung-Galaxy-Ace-Plus.html' · '/Samsung-Galaxy-A20.html' · '/Samsung-Galaxy-Chat.html' ·
'/Samsung-Galaxy-Gio.html' · '/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·
'/Samsung-Galaxy-Tab-A-10.5WIFI.html' · '/Samsung-Galaxy-Alpha.html' ·
'/Samsung-Galaxy-S3-Slim.html' · '/Samsung-Galaxy-S4-zoom.html' ·
'/Samsung-Galaxy-Xcover-2.html' · '/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·
'/Samsung-Galaxy-Tab-A8-LTE-2021.html' · '/Samsung-Galaxy-A8-2018.html' ·
'/Samsung-Galaxy-Tab4-8.0-wifi.html' · '/Samsung-Galaxy-M33-5G.html' ·
'/Samsung-Galaxy-A50.html' · '/Samsung-Galaxy-E7.html' · '/Samsung-Galaxy-S6.html' ·
'/Samsung-Galaxy-S20-FE.html' · '/Samsung-Galaxy-Tab-S4-WIFI.html' ·
'/Samsung-Galaxy-S7.html' · '/Samsung-Galaxy-Note-5-Exynos.html' ·
'/Samsung-Galaxy-TabPRO-12.2-LTE.html' · '/Samsung-Galaxy-S4-Active.html' ·
'/Samsung-Galaxy-Tab-Active-3.html' · '/Samsung-Galaxy-Tab-S3-9.7.html' ·
'/Samsung-Galaxy-S6-edge.html' · '/Samsung-Galaxy-Note-4-Exynos.html' ·
'/Samsung-Galaxy-Round.html' · '/Samsung-Galaxy-Note-20-Ultra-5G.html' ·
'/Samsung-ATIV-Q.html' · '/Samsung-ATIV-Smart-PC-PRO.html' ·
'/Samsung-Galaxy-S22-Ultra12-128GB.html' · '/Samsung-Galaxy-Z-Flip-5G.html' ·
'/Samsung-Galaxy-Z-Flip.html' · '/Samsung-Galaxy-Tab-S8-Ultra-5G.html' ·
'/Samsung-Galaxy-S21-Ultra-16-512GB.html' · '/Samsung-Galaxy-S10-Plus-Ram-12GB.html' ·
'/Samsung-Galaxy-Z-Fold-3.html' · '/Samsung-Galaxy-Z-Fold4.html' ·
'/Samsung-Galaxy-Z-Fold-2-5G.html'

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

full_links

'http://specphone.com/Samsung-Galaxy-M13.html' ·
'http://specphone.com/Samsung-Galaxy-A23.html' ·
'http://specphone.com/Samsung-Galaxy-A13.html' ·
'http://specphone.com/Samsung-Galaxy-M32-5G.html' ·
'http://specphone.com/Samsung-Galaxy-A12-Nacho.html' ·
'http://specphone.com/Samsung-Galaxy-Pocket-Neo.html' ·
'http://specphone.com/Samsung-Galaxy-Young.html' ·
'http://specphone.com/Samsung-Galaxy-J1-Mini.html' ·
'http://specphone.com/Samsung-Galaxy-A01-Core-1-16GB.html' ·
'http://specphone.com/Samsung-Galaxy-V-PLUS.html' ·
'http://specphone.com/Samsung-Galaxy-Young-2.html' ·
'http://specphone.com/Samsung-Galaxy-M02.html'

'http://specphone.com/Samsung-Galaxy-M02.html' ·
'http://specphone.com/Samsung-Galaxy-A11.html' ·
'http://specphone.com/Samsung-Galaxy-J2-Pro-2018.html' ·
'http://specphone.com/Samsung-Galaxy-A12-2021.html' ·
'http://specphone.com/Samsung-Galaxy-A21s-3-32GB.html' ·
'http://specphone.com/Samsung-Galaxy-J5.html' ·
'http://specphone.com/Samsung-Galaxy-J4.html' ·
'http://specphone.com/Samsung-Galaxy-Core-2-Duos.html' ·
'http://specphone.com/Samsung-Galaxy-Ace-Plus.html' ·
'http://specphone.com/Samsung-Galaxy-A20.html' ·
'http://specphone.com/Samsung-Galaxy-Chat.html' ·
'http://specphone.com/Samsung-Galaxy-Gio.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-A-10.5WIFI.html' ·
'http://specphone.com/Samsung-Galaxy-Alpha.html' ·
'http://specphone.com/Samsung-Galaxy-S3-Slim.html' ·
'http://specphone.com/Samsung-Galaxy-S4-zoom.html' ·
'http://specphone.com/Samsung-Galaxy-Xcover-2.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-A8-LTE-2021.html' ·
'http://specphone.com/Samsung-Galaxy-A8-2018.html' ·
'http://specphone.com/Samsung-Galaxy-Tab4-8.0-wifi.html' ·
'http://specphone.com/Samsung-Galaxy-M33-5G.html' ·
'http://specphone.com/Samsung-Galaxy-A50.html' ·
'http://specphone.com/Samsung-Galaxy-E7.html' ·
'http://specphone.com/Samsung-Galaxy-S6.html' ·
'http://specphone.com/Samsung-Galaxy-S20-FE.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-S4-WIFI.html' ·
'http://specphone.com/Samsung-Galaxy-S7.html' ·
'http://specphone.com/Samsung-Galaxy-Note-5-Exynos.html' ·
'http://specphone.com/Samsung-Galaxy-TabPRO-12.2-LTE.html' ·
'http://specphone.com/Samsung-Galaxy-S4-Active.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-Active-3.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-S3-9.7.html' ·
'http://specphone.com/Samsung-Galaxy-S6-edge.html' ·
'http://specphone.com/Samsung-Galaxy-Note-4-Exynos.html' ·
'http://specphone.com/Samsung-Galaxy-Round.html' ·
'http://specphone.com/Samsung-Galaxy-Note-20-Ultra-5G.html' ·
'http://specphone.com/Samsung-ATIV-Q.html' ·
'http://specphone.com/Samsung-ATIV-Smart-PC-PRO.html' ·
'http://specphone.com/Samsung-Galaxy-S22-Ultra12-128GB.html' ·
'http://specphone.com/Samsung-Galaxy-Z-Flip-5G.html' ·
'http://specphone.com/Samsung-Galaxy-Z-Flip.html' ·
'http://specphone.com/Samsung-Galaxy-Tab-S8-Ultra-5G.html' ·
'http://specphone.com/Samsung-Galaxy-S21-Ultra-16-512GB.html' ·
'http://specphone.com/Samsung-Galaxy-S10-Plus-Ram-12GB.html' ·
'http://specphone.com/Samsung-Galaxy-Z-Fold-3.html' ·
'http://specphone.com/Samsung-Galaxy-Z-Fold4.html' ·
'http://specphone.com/Samsung-Galaxy-Z-Fold-2-5G.html'

```r
result <- data.frame()

for (links in full_links[1:5]) {
    ss_topic <- links %>%
        read_html() %>%
        html_nodes("div.topic") %>%
        html_text2()

    ss_detail <- links %>%
        read_html() %>%
        html_nodes("div.detail") %>%
        html_text2()

    tmp <- data.frame(attribute = ss_topic,value = ss_detail)
    result <-  bind_rows(result,tmp)
    print("Progress..... ")
}

print(result)
```

```r
write_csv(result,"ss_phone.csv")
```

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

```
    attribute                                           value
1                                                       2565
2
3                              165.40 x 76.90 x 8.40    .
4                                                        192
5          Glass front, plastic back, plastic frame
6       SIM           2         (nano sim, nano sim)
```