

Mini Project 01 - IMDB Web Scrapping

```
# create markdown cell (clrt + m)
```

```
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"
```

```
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. The Lord of the Rings: The Return of the King (2003)' · '5. Schindler\'s List (1993)' ·  
'6. The Godfather Part II (1974)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·  
'10. The Lord of the Rings: The Two Towers (2002)'
```

```
# load rating
```

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

```
# load vote & gross
```

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

```
# build a dataset
```

```
df <- data.frame(  
  Title = titles,  
  Rating = ratings,  
  Num_vote = numvote  
)
```

```
head(df)
```

A data.frame: 6 × 3

	Title	Rating	Num_vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,672,027 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,851,869 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,644,857 Gross: \$534.86M Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,841,485 Gross: \$377.85M Top 250: #7
5	5. Schindler's List (1993)	9.0	Votes: 1,352,658 Gross: \$96.90M Top 250: #6
6	6. The Godfather Part II (1974)	9.0	Votes: 1,268,013 Gross: \$57.30M Top 250: #4

Mini Project 02 - Specphone Database

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

```
url <-read_html("https://specphone.com/Samsung-Galaxy-Note-20-Ultra-5G.html")
```

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

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

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

A data.frame: 34 × 2

att	value
<chr>	<chr>
วันเปิดตัว	สิงหาคม 2563
วันวางจำหน่าย	สิงหาคม 2563, วางจำหน่ายแล้ว
ขนาด	164.80 x 77.20 x 8.10 มม.
น้ำหนัก	ไม่รองรับ
วัสดุ	Gorilla Glass 7, aluminum frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	5G,LTE,HSPA+
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	850/900/1900/2100/2600
ความเร็ว	5G,LTE,HSPA+
ประเภท	Dynamic AMOLED 2X Infinity-O Display
ขนาดหน้าจอ	6.90 นิ้ว
ความละเอียด	3000 x 1440 pixels
ฟีเจอร์เพิ่มเติม	หน้าจอ refresh rate 120Hz , HDR10+ รองรับ Samsung DeX จอกันรอย Gorilla Glass 7 กันน้ำกันฝุ่น มาตรฐาน IP68
ระบบปฏิบัติการ	Android 10
ชิปประมวลผล	Samsung Exynos 990 2.73 GHz
ชิปกราฟิก	Mali G77 MP11
หน่วยความจำ	12 GB
ความจุ	256/512 GB
Memory Card	3.2, Type- (256)
กล้องหลัก	ตัวที่ 1: 108 MP, f/1.8, 26mm (wide), 1/1.33 ตัวที่ 2: 12 MP, f/3.0, 103mm (periscope telephoto), 1.0µm, PDAF, OIS, 5x optical zoom, 50x hybrid zoom ตัวที่ 3: 12 MP, f/2.2, 13mm (ultrawide), 1/2.55
ความละเอียดวิดีโอ	4K@30/60fps, 1080p@30fps
กล้องหน้า	ตัวที่ 1: 10 MP, f/2.2, 26mm (wide), 1/3.2
Bluetooth	5.0, A2DP, LE, aptX
Wi-Fi	802.11 a/b/g/n/ac/6
USB	addnew
GPS	GPS, Galileo, Glonass, Be
NFC	รองรับ
ความจุ	4,500 mAh
ประเภท	Li - Polymer
Wireless Charging	รองรับ
Fast Charging	รองรับ (SuperFast Charge,Fast Qi 15W)

```
# all samsung
```

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

```
links <- samsung_url %>%  
  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_link <- paste0("https://specphone.com",links)
```

```
full_link
```

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

```

result <- data.frame()

for (links in full_link[1:10]){
  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)

```

```

[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

```

```

write_csv2(result,"result_ss.csv")

```