

Mini Project 01 - IMDB web scraping

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

```
url <- "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()
```

```
title[1:10]
```

ERROR: Error in title[1:10]: object of type 'closure' is not subsettable

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

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

```
# build a dataset
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: 2,709,329 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,881,474 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,682,358 Gross: \$534.86M Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,368,887 Gross: \$96.90M Top 250: #6
5	5. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,864,698 Gross: \$377.85M Top 250: #7
6	6. The Godfather Part II (1974)	9.0	Votes: 1,284,571 Gross: \$57.30M Top 250: #4

MiniProject - Specphone web scraping

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

```
url <- read_html("https://specphone.com/OnePlus-11.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: 32 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	กุมภาพันธ์ 2566
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	163.10 x 74.10 x 8.50 มม.
น้ำหนัก	205 กรัม
วัสดุ	Glass front, plastic frame, plastic back
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA, 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, LTE-A (CA), 5G
ประเภท	LTPO3 Fluid AMOLED
ขนาดหน้าจอ	6.70 นิ้ว
ความละเอียด	1440 x 3216 pixels
ระบบปฏิบัติการ	Android 13
ชิปประมวลผล	Qualcomm Snapdragon 8 Gen 2 SM8550 3.2 GHz
ชิปกราฟิก	Adreno 740
หน่วยความจำ	8 GB
ความจุ	128 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, 24mm (wide), 1/1.56 ตัวที่ 2: 32 MP, f/2.0, 48mm (telephoto), 1/2.74 ตัวที่ 3: 48 MP, f/2.2, 115°, (ultrawide), 1/2.0
ความละเอียดวิดีโอ	8K@24fps, 4K@30/60fps, 1080p@30/60/240fps, Auto HDR, gyro-EIS
กล้องหน้า	ตัวที่ 1: 16 MP, f/2.5, 25mm (wide), 1.0µm
Bluetooth	5.3, A2DP, LE, aptX HD
Wi-Fi	802.11 a/b/g/n/ac/6, dual
USB	Type-C
GPS	dual-band A-GPS, GLONASS,
NFC	รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (80W)

```
# all one plus smartphone
oneplus_url <- read_html("https://specphone.com/brand/OnePlus")
```

```
links <- oneplus_url %>%
  html_nodes("li.mobile-brand-item a") %>% ## เว้นวรรคแล้ว a แปลว่า จงหา ตัว a ใน mob:
  html_attr("href")
```

links

```
'/OnePlus-Pad.html' · '/OnePlus-11R.html' · '/OnePlus-Nord-N20-5G.html' · '/OnePlus-10R.html' ·
'/OnePlus-Nord-N200-5G.html' · '/OnePlus-9-Pro.html' · '/OnePlus-9.html' · '/OnePlus-Nord-N100.html' ·
'/OnePlus-Nord-CE-2-Lite-5G.html' · '/OnePlus-Nord-N10-5G.html' · '/OnePlus-2-16-GB.html' ·
'/OnePlus-Nord-CE-5G8-128GB.html' · '/OnePlus-2-64-GB.html' · '/OnePlus-Nord-2T.html' ·
'/OnePlus-Nord-CE-2-5G.html' · '/OnePlus-Nord-5G-8-128GB.html' · '/OnePlus-Nord-CE-5G12-256GB.html' ·
'/OnePlus-3T.html' · '/OnePlus-3.html' · '/OnePlus-One.html' · '/OnePlus-Nord-5G.html' ·
'/OnePlus-Nord-2-5G8-128GB.html' · '/OnePlus-6T.html' · '/OnePlus-7-Pro-Ram-6GB.html' ·
'/OnePlus-10T8-128GB.html' · '/OnePlus-8T-8GB.html' · '/OnePlus-7-Pro-Ram-8GB.html' ·
'/OnePlus-7T-Pro-Ram-8GB.html' · '/OnePlus-10T12-256GB.html' · '/OnePlus-7-Pro-Ram-12GB.html' ·
'/OnePlus-8-12-256G.html' · '/OnePlus-11.html' · '/OnePlus-8T-12GB.html' · '/OnePlus-8-Pro.html'
```

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

```
full_links[1:10]
```

```
'https://specphone.com/OnePlus-Pad.html' · 'https://specphone.com/OnePlus-11R.html' ·
'https://specphone.com/OnePlus-Nord-N20-5G.html' · 'https://specphone.com/OnePlus-10R.html' ·
'https://specphone.com/OnePlus-Nord-N200-5G.html' · 'https://specphone.com/OnePlus-9-Pro.html' ·
'https://specphone.com/OnePlus-9.html' · 'https://specphone.com/OnePlus-Nord-N100.html' ·
'https://specphone.com/OnePlus-Nord-CE-2-Lite-5G.html' · 'https://specphone.com/OnePlus-Nord-N10-5G.html'
```

```

result <- data.frame()

for (link in full_links[1:10]){
  op_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()
  op_detail <- link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()
  tmp <- data.frame(attribute = op_topic,
                    value = op_detail)

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

```

```

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

```

```
print(result)
```

	attribute
1	วันเปิดตัว
2	วันวางจำหน่าย
3	ขนาด
4	น้ำหนัก
5	วัสดุ
6	SIM
7	Technology
8	2G
9	3G
10	4G
11	5G
12	ความเร็ว
13	ประเภท
14	ขนาดหน้าจอ
15	ความละเอียด
16	ระบบปฏิบัติการ

17 ขีปประมวลผล
18 ขีปกราฟิก

```
write_csv(result, "result-op.csv")
```

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

	attribute	value
1	วันเปิดตัว	กุมภาพันธ์ 2566
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	258.00 x 189.00 x 6.50 มม.