

Mini Project 01 - IMDB Web Scraping

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

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,des
#read HTML
imdb <- read_html(url)
```

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

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

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

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

A data.frame: 6 × 3

	title	rating	vote
	<chr>	<chr>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,693,344 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,868,934 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,667,168 Gross: \$534.86M Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,362,192 Gross: \$96.90M Top 250: #6
5	5. The Godfather Part II (1974)	9.0	Votes: 1,277,922 Gross: \$57.30M Top 250: #4
6	6. 12 Angry Men (1957)	9.0	Votes: 795,787 Gross: \$4.36M Top 250: #5

Mini Project 02 - Specphone Phone Databased

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

```
url <- read_html("https://specphone.com/0PP0-A78-5G-8-128GB.html")
```

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

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

A data.frame: 32 × 2

attributes	value
<chr>	<chr>
วันเปิดตัว	มกราคม 2566
วันวางจำหน่าย	วางจำหน่ายแล้ว
ขนาด	163.80 x 75.10 x 8.00 มม.
น้ำหนัก	188 กรัม
วัสดุ	Glass front, plastic frame, plastic back
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	IPS LCD
ขนาดหน้าจอ	6.56 นิ้ว
ความละเอียด	720 x 1612 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	MediaTek Dimensity 700 5G MT6833 2.2 GHz
ชิปกราฟิก	Mali-G57 MC2
หน่วยความจำ	8 GB
ความจุ	128 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, 27mm (wide), PDAF ตัวที่ 2: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 8 MP, f/2.0, 27mm (wide)
Bluetooth	5.3, A2DP, LE, aptX
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	GPS, GALILEO, GLONASS, BD
NFC	รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (33W)

All OPPO Phone

```
oppo_url <- read_html("https://specphone.com/brand/OPP0")
```

```
links <- oppo_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[1:5]){  
  oppo_topic <- link %>%  
    read_html() %>%  
    html_nodes("div.topic") %>%  
    html_text2()  
  oppo_detail <- link %>%  
    read_html() %>%  
    html_nodes("div.detail") %>%  
    html_text2()  
  tmp <- data.frame(attribute = oppo_topic,  
                    value = oppo_detail)  
  result <- bind_rows(result, tmp)  
  print("Progress ...")  
}
```

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

```
write_csv(result, "result_oppo_phone.csv")
```

```
head(result)
```

A data.frame: 6 × 2

	attribute	value
	<chr>	<chr>
1	วันเปิดตัว	ธันวาคม 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	166.20 x 75.20 x 7.50 มม.
4	น้ำหนัก	191 กรัม
5	วัสดุ	Glass front (Gorilla Glass Victus) - folded, plastic front - unfolded, glass back (Gorilla Glass Victus) or eco leather back, aluminum frame
6	SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)