

# Mini Project 01 - IMDB Web Scrapping

```
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 <- "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)
```

imdb

```
{html_document}
<html xmlns:og="http://ogp.me/ns#" xmlns:fb="http://www.facebook.com/2008/fbml">
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 ...
[2] <body id="styleguide-v2" class="fixed">\n                <img height="1" widt ...
```

```
# movie titles
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)'

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

ratings[1:10]

'9.3' · '9.2' · '9.0' · '9.0' · '9.0' · '9.0' · '9.0' · '8.9' · '8.8' · '8.8'

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

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

```
df[1:15, ]
```

A data.frame: 15 × 2

	attribute	detail
	<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)
7	Technology	HSPA, LTE-A (CA), 5G
8	2G	850/900/1800/1900
9	3G	850/900/1900/2100
10	4G	850/900/1900/2100/2600
11	5G	2100/2600/3500/4700
12	ความเร็ว	HSPA, LTE-A (CA), 5G
13	ประเภท	Foldable AMOLED
14	ขนาดหน้าจอ	6.80 นิ้ว
15	ความละเอียด	1080 x 2520 pixels

## Mini Project 02 - Specphone Phone Database

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

```
url <- read_html("https://specphone.com/OPP0-Find-N2-Flip.html")
```

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

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

```
df <- data.frame(
  attribute = att,
  detail = value
)
```

```
df
```

attribute	detail
<chr>	<chr>
วันเปิดตัว	ธันวาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	166.20 x 75.20 x 7.50 มม.
น้ำหนัก	191 กรัม
วัสดุ	Glass front (Gorilla Glass Victus) - folded, plastic front - unfolded, glass back (Gorilla Glass Victus) or eco leather back, aluminum frame
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	Foldable AMOLED
ขนาดหน้าจอ	6.80 นิ้ว
ความละเอียด	1080 x 2520 pixels
ระบบปฏิบัติการ	Android 13
ชิปประมวลผล	MediaTek Dimensity 9000+ null 3.35 GHz
ชิปกราฟิก	Mali-G710 MC10
หน่วยความจำ	12 GB
ความจุ	256 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, 23mm (wide), PDAF ตัวที่ 2: 8 MP, f/2.2, 112°, (ultrawide)
ความละเอียดวิดีโอ	4K@30/60fps, 1080p@30/60/240fps
กล้องหน้า	ตัวที่ 1: 32 MP, f/2.4, 22mm (wide), AF
Bluetooth	5.3, A2DP, LE, aptX HD
Wi-Fi	802.11 a/b/g/n/ac/6, tri-
USB	Type-C
GPS	GPS (L1+L5), GLONASS (G1)
NFC	รองรับ
ความจุ	4,300 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (44W)

```
# all oppo smartphone
oppo_url <- read_html("https://specphone.com/brand/OPPO")
```

```
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[1:10]) {
  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(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	ชิปประมวลผล

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

```
      attribute
1   วันเปิดตัว
2 วันวางจำหน่าย
3   ขนาด
4   น้ำหนัก
5   วัสดุ
6   SIM
```

```
1
2
3
4
5 Glass front (Gorilla Glass Victus) - folded, plastic front - unfolded, glass back (Gor
6
```

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