

## Mini Project 01 - IMDB web scraping

```
library(tidyverse)
library(rvest) # scrap data from internet
```

```
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"
```

```
# READ HTML
```

```
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" width .
```

```
# movie title

titles <- imdb %>%
  html_nodes("h3.list-item-header") %>% #node == tag ตัวนี้
  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() %>%
  as.numeric()
```

```
ratings
```

```
9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8 · 8.8 · 8.8 · 8.8 · 8.7 · 8.7 · 8.7 · 8.7 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 ·
8.6 · 8.6 · 8.6 · 8.6 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5 · 8.5
```

```
# number of votes

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

```
# build 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: 2,676,557   Gross: \$28.34M   Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,855,349   Gross: \$134.97M   Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,649,867   Gross: \$534.86M   Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,844,500   Gross: \$377.85M   Top 250: #7
5	5. Schindler's List (1993)	9.0	Votes: 1,354,767   Gross: \$96.90M   Top 250: #6
6	6. The Godfather Part II (1974)	9.0	Votes: 1,270,156   Gross: \$57.30M   Top 250: #4

## Project 02 - Specphone Phone Database

```
library(tidyverse)
library(rvest) # scrap data from internet
```

```
url <- read_html("https://specphone.com/Apple-iPhone-14-Pro-Max.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: 33 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	กันยายน 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	160.70 x 77.60 x 7.90 มม.
น้ำหนัก	240 กรัม
วัสดุ	Glass front (Gorilla Glass), glass back (Gorilla Glass), stainless steel frame
SIM	รองรับ 1 ซิมการ์ด (esim,nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A, 5G, EV-DO Rev.A 3.1 Mbps
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	2100/2600/3500/4700
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A, 5G, EV-DO Rev.A 3.1 Mbps
ประเภท	LTPO Super Retina XDR OLED
ขนาดหน้าจอ	6.70 นิ้ว
ความละเอียด	1290 x 2796 pixels
ระบบปฏิบัติการ	iOS 16
ชิปประมวลผล	Apple null A16 Bionic 3.1 GHz
ชิปกราฟิก	Apple GPU (5-core graphics)
หน่วยความจำ	6 GB
ความจุ	128/256/512/1024 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 48 MP, f/1.8, 24mm (wide), 1.22µm, dual pixel PDAF, sensor-shift OIS ตัวที่ 2: 12 MP, f/2.8, 77mm (telephoto), PDAF, OIS, 3x optical zoom ตัวที่ 3: 12 MP, f/2.2, 13mm, 120° (ultrawide), 1.4µm, dual pixel PDAF ตัวที่ 4: TOF 3D LiDAR scanner (depth)
ความละเอียดวิดีโอ	4K@24/25/30/60fps, 1080p@25/30/60/120/240fps, HDR, Dolby Vision HDR (up to 60fps), Cinematic mode (4K@30fps), stereo sound rec.
กล้องหน้า	ตัวที่ 1: 12 MP, f/1.9, 23mm (wide), 1/3.6 ตัวที่ 2: SL 3D, (depth/biometrics sensor)
Bluetooth	5.3, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac/6, dual
USB	Lightning
GPS	dual-band A-GPS, GLONASS,
NFC	รองรับ
ความจุ	0 mAh
ประเภท	Non-removable Li-Ion
Wireless Charging	รองรับ
Fast Charging	รองรับ ()

```
# all iphone

ios_url <- read_html("https://specphone.com/brand/Apple")
```

```
# links to all ios phone

links <- ios_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]) {
  ios_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

  ios_detail <- link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()

  tmp <- data.frame(attribute = ios_topic,
                    value = ios_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
```

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

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

	attribute	value
1	วันเปิดตัว	พฤศจิกายน 2555
2	วันวางจำหน่าย	พฤษภาคม 2556, สินค้าจำหน่ายหมดแล้ว
3	ขนาด	200.00 x 134.70 x 7.20 มม.
4	น้ำหนัก	308 กรัม
5	วัสดุ	Aluminium, Plastic
6	SIM	รองรับ 1 ซิมการ์ด