

MINI PROJECT 01 - IMDB WEB SCRAPING

```
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"
```

```
# 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.lister-item-header") %>%
  html_text2()
```

```
titles[1:10]
```

```
'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' ·
'4. Schindler's List (1993)' · '5. The Lord of the Rings: The Return of the King (2003)' ·
'6. The Godfather Part II (1974)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' ·
'9. The Lord of the Rings: The Fellowship of the Ring (2001)' · '10. Inception (2010)'
```

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

```
ratings[1:10]
```

```
9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8
```

```
# 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: 2,703,533 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,877,126 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,677,342 Gross: \$534.86M Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,366,483 Gross: \$96.90M Top 250: #6
5	5. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,861,475 Gross: \$377.85M Top 250: #7
6	6. The Godfather Part II (1974)	9.0	Votes: 1,282,241 Gross: \$57.30M Top 250: #4

MINI PROJECT 02 - SpecPhone PHONE DATABASE

```
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
```

```
✓ tidy  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 <- 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
Foot	

```
# all apple phone
apple_url <- read_html("https://specphone.com/brand/Apple")
```

```
# link to all apple phone
links <- apple_url %>%
  html_nodes("li.mobile-brand-item a") %>%
  html_attr("href")
```

```
links
```

```

'/Apple-iPad-Mini-Wifi.html' · '/Apple-iPhone-3GS-8GB.html' · '/Apple-iPad-Mini-2-Wifi.html' ·
'/Apple-iPad-10.2-WiFi-2020.html' · '/Apple-iPad-10.2.html' · '/Apple-iPad-10.2-WiFi2021.html' ·
'/Apple-iPad-9.7-2018-WiFi.html' · '/Apple-iPhone-4-8GB.html' · '/Apple-iPhone-6s-Plus.html' ·
'/Apple-iPad-4-Wifi.html' · '/Apple-iPhone-6.html' · '/Apple-iPad-Air-2-WiFi.html' ·
'/Apple-iPad-Wi-Fi-16GB.html' · '/Apple-iPad-Mini-Wifi+Cellular.html' · '/Apple-iPad-Mini-3-WiFi.html' ·
'/Apple-iPad-Air-5-2022-WiFi.html' · '/Apple-iPhone-5.html' · '/Apple-iPad-2-Wi-Fi-64GB.html' ·
'/Apple-iPhone-XR.html' · '/Apple-iPhone-4.html' · '/Apple-iPad-Air-2019-Cellular.html' ·
'/Apple-iPad-9.7-WiFi.html' · '/Apple-iPad-2-3G-32GB.html' · '/Apple-iPad-3G-64GB.html' ·
'/Apple-iPad-mini-6-Cellular.html' · '/Apple-iPad-Air-2020-Cellular.html' · '/Apple-iPad-Pro-10.5-WiFi.html' ·
'/Apple-iPhone-14-Pro.html' · '/Apple-iPhone-13-Pro-Max.html' · '/Apple-iPad-Pro-12.9-Cellular-2021.html' ·
'/Apple-iPhone-14-Pro-Max.html'
```

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

```

result <- data.frame()

for (link in full_links[1:5]) {

  apple_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

  apple_detail <- link %>%
```

```

read_html() %>%
html_nodes("div.detail") %>%
html_text2()

tmp <- data.frame(attribute = apple_topic,
                  value = apple_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	ฟีเจอร์เพิ่มเติม
17	ระบบปฏิบัติการ
18	ชิปประมวลผล
19	ชิปกราฟิก

```

# write csv
write_csv(result, "result_apple_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