

Mini Project 01 - IMDB Database

```
library(tidyverse)
library(rvest) #scrape 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.lister-item-header") %>%
  html_text2()
```

```
titles[1:5]
```

```
'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' ·
'4. The Godfather Part II (1974)' · '5. Schindler's List (1993)'
```

```
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
```

Mini Project 02 - Specphone Database

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

```
url <- read_html("https://specphone.com/Asus-ROG-Phone-6-Diablo-Immortal-Edition.
```

```
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	173.00 x 77.00 x 10.30 มม.
น้ำหนัก	239 กรัม
วัสดุ	Glass front (Gorilla Glass Victus), glass back (Gorilla Glass 3), aluminum frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	Up to HSPA 42.2/5.76 Mbps, LTE-A (6CA) Cat20 2024/210 Mbps
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	2100/2600/3500/4700
ความเร็ว	Up to HSPA 42.2/5.76 Mbps, LTE-A (6CA) Cat20 2024/210 Mbps
ประเภท	AMOLED
ขนาดหน้าจอ	6.78 นิ้ว
ความละเอียด	1080 x 2448 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Qualcomm Snapdragon 8+ Gen 1 SM8475 3.19 GHz
ชิปกราฟิก	Adreno 730
หน่วยความจำ	16 GB
ความจุ	512 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.9, 24mm (wide), 1/1.56 ตัวที่ 2: 13 MP, f/2.2, 13mm (ultrawide) ตัวที่ 3: 5 MP, (macro)
ความละเอียดวิดีโอ	8K@24fps, 4K@30/60fps, 4K@120fps, 1080p@30/60/120/240fps, 720p@480fps; gyro-EIS, HDR10+
กล้องหน้า	ตัวที่ 1: 12 MP, 28mm (wide)
Bluetooth	5.2, A2DP, LE, aptX HD, aptX Adaptive
Wi-Fi	802.11 a/b/g/n/ac/6e, tri
USB	Type-C
GPS	A-GPS. Up to dual-band: G
NFC	รองรับ
ความจุ	6,000 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (65W)

```
# All samsung smartphone
samsung_url <- read_html("https://specphone.com/brand/Samsung")
```

```
# link to all samsung smartphone
links <- samsung_url %>%
  html_nodes("li.mobile-brand-item a") %>%
  html_attr("href")
links
```

```
'/Samsung-Galaxy-M13.html' · '/Samsung-Galaxy-A23.html' · '/Samsung-Galaxy-A13.html' ·
'/Samsung-Galaxy-M32-5G.html' · '/Samsung-Galaxy-A12-Nacho.html' · '/Samsung-Galaxy-Pocket-Neo.html' ·
'/Samsung-Galaxy-Young.html' · '/Samsung-Galaxy-J1-Mini.html' · '/Samsung-Galaxy-A01-Core-1-16GB.html' ·
'/Samsung-Galaxy-V-PLUS.html' · '/Samsung-Galaxy-Young-2.html' · '/Samsung-Galaxy-M02.html' ·
'/Samsung-Galaxy-A11.html' · '/Samsung-Galaxy-J2-Pro-2018.html' · '/Samsung-Galaxy-A12-2021.html' ·
'/Samsung-Galaxy-A21s-3-32GB.html' · '/Samsung-Galaxy-J5.html' · '/Samsung-Galaxy-J4.html' ·
'/Samsung-Galaxy-Core-2-Duos.html' · '/Samsung-Galaxy-Ace-Plus.html' · '/Samsung-Galaxy-A20.html' ·
'/Samsung-Galaxy-Chat.html' · '/Samsung-Galaxy-Gio.html' · '/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·
'/Samsung-Galaxy-Tab-A-10.5WIFI.html' · '/Samsung-Galaxy-Alpha.html' · '/Samsung-Galaxy-S3-Slim.html' ·
'/Samsung-Galaxy-S4-zoom.html' · '/Samsung-Galaxy-Xcover-2.html' · '/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·
'/Samsung-Galaxy-Tab-A8-LTE-2021.html' · '/Samsung-Galaxy-A8-2018.html' ·
'/Samsung-Galaxy-Tab4-8.0-wifi.html' · '/Samsung-Galaxy-M33-5G.html' · '/Samsung-Galaxy-A50.html' ·
'/Samsung-Galaxy-E7.html' · '/Samsung-Galaxy-S6.html' · '/Samsung-Galaxy-S20-FE.html' ·
'/Samsung-Galaxy-Tab-S4-WIFI.html' · '/Samsung-Galaxy-S7.html' · '/Samsung-Galaxy-Note-5-Exynos.html' ·
'/Samsung-Galaxy-TabPRO-12.2-LTE.html' · '/Samsung-Galaxy-S4-Active.html' ·
'/Samsung-Galaxy-Tab-Active-3.html' · '/Samsung-Galaxy-Tab-S3-9.7.html' · '/Samsung-Galaxy-S6-edge.html' ·
'/Samsung-Galaxy-Note-4-Exynos.html' · '/Samsung-Galaxy-Round.html' ·
'/Samsung-Galaxy-Note-20-Ultra-5G.html' · '/Samsung-ATIV-Q.html' · '/Samsung-ATIV-Smart-PC-PRO.html' ·
'/Samsung-Galaxy-S23-Ultra-5G8-256GB.html' · '/Samsung-Galaxy-S22-Ultra12-128GB.html' ·
'/Samsung-Galaxy-Z-Flip-5G.html' · '/Samsung-Galaxy-Z-Flip.html' · '/Samsung-Galaxy-Tab-S8-Ultra-5G.html' ·
'/Samsung-Galaxy-S21-Ultra-16-512GB.html' · '/Samsung-Galaxy-S23-Ultra-5G.html' ·
'/Samsung-Galaxy-S10-Plus-Ram-12GB.html' · '/Samsung-Galaxy-Z-Fold-3.html'
```

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

```

result <- data.frame()

for (link in full_links[1:10]) {
  ss_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

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

  tmp <- data.frame(attribute = ss_topic,
                    value = ss_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
9   3G

```

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

	attribute	value
1	วันเปิดตัว	มิถุนายน 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	165.40 x 76.90 x 8.40 มม.
4	น้ำหนัก	192 กรัม

5	วัสดุ	Glass front, plastic back, plastic frame
6	SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)

```
# write csv  
write_csv(result, "result_ss_phone.csv")
```