

Mini Project 01 - IMDB web scraping

```
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/fbm"
[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 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. 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[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: 2,679,311 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,857,481 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,652,669 Gross: \$534.86M Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,846,445 Gross: \$377.85M Top 250: #7
5	5. Schindler's List (1993)	9.0	Votes: 1,355,902 Gross: \$96.90M Top 250: #6
6	6. The Godfather Part II (1974)	9.0	Votes: 1,271,479 Gross: \$57.30M Top 250: #4

Mini Project 02 - Specphone Phone Database

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

```
url <- read_html("https://specphone.com/ZTE-nubia-Red-Magic-8-Pro-.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: 32 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ธันวาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	164.00 x 76.40 x 8.90 มม.
น้ำหนัก	230 กรัม
วัสดุ	Glass front, glass back, aluminum frame
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	AMOLED
ขนาดหน้าจอ	6.80 นิ้ว
ความละเอียด	1116 x 2480 pixels
ระบบปฏิบัติการ	Android 13
ชิปประมวลผล	Qualcomm Snapdragon 8 Gen 2 SM8550 3.2 GHz
ชิปกราฟิก	Adreno 740
หน่วยความจำ	12 GB
ความจุ	256 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), 1/1.57 ตัวที่ 2: 8 MP, f/2.2, 120°, 13mm (ultrawide), 1/4.0 ตัวที่ 3: 2 MP, f/2.4, (macro)
ความละเอียดวิดีโอ	8K@30fps, 4K@30/60fps, 1080p@30/60/120/240fps
กล้องหน้า	ตัวที่ 1: 16 MP, (wide), under display
Bluetooth	5.3, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac/6e, dua
USB	Type-C
GPS	GPS (L1+L5), GLONASS, BDS
NFC	รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (165W)

```
# All ZTE Smartphones
zte_url <- read_html("https://specphone.com/brand/ZTE")
```

```
links <- zte_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]) {  
  zte_topic <- link %>%  
    read_html() %>%  
    html_nodes("div.topic") %>%  
    html_text2()  
  
  zte_detail <- link %>%  
    read_html() %>%  
    html_nodes("div.detail") %>%  
    html_text2()  
  
  tmp <- data.frame(attribute = zte_topic,  
                    value = zte_detail)  
  result <- bind_rows(result, tmp)  
  print("Progress ...")  
}  
  
print(result)
```

```
[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  
10     4G  
11     5G  
12     ความเร็ว  
13     ประเภท  
14     ขนาดหน้าจอ
```

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

	attribute	value
1	วันเปิดตัว	ธันวาคม 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	164.00 x 76.40 x 8.90 มม.
4	น้ำหนัก	230 กรัม
5	วัสดุ	Glass front, glass back, aluminum frame
6	SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)

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