

Mini Project 01 - IMDb web scarping

```
library(tidyverse)
library(rvest) #web scarping
```

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
imdb <- read_html(url)
```

```
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. The Godfather Part II (1974)' ·
'6. Schindler's List (1993)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' ·
'9. The Lord of the Rings: The Fellowship of the Ring (2001)' · '10. Fight Club (1999)'
```

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

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

```
#build a dataset
df <- data.frame(
  title = titles,
  rating = rating,
  vote = vote
)
head(df)
```

A data.frame: 6 × 3

	title	rating	vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,682,373 Gross: \$28.34M Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,859,744 Gross: \$134.97M Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,655,734 Gross: \$534.86M Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,848,438 Gross: \$377.85M Top 250: #7
5	5. The Godfather Part II (1974)	9.0	Votes: 1,272,930 Gross: \$57.30M Top 250: #4
6	6. Schindler's List (1993)	9.0	Votes: 1,357,234 Gross: \$96.90M Top 250: #6

Mini Project 02 - Specphone Phone database

```
library(tidyverse)
library(rvest) #web scarping
```

```
url <- "https://specphone.com/Samsung-Galaxy-A30.html"
specphone <- read_html(url)
```

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

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

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

A data.frame: 6 × 2

	att	value
	<chr>	<chr>
1	วันเปิดตัว	เมษายน 2562
2	วันวางจำหน่าย	เมษายน 2562, วางจำหน่ายแล้ว
3	ขนาด	157.50 x 74.70 x 7.70 มม.
4	น้ำหนัก	ไม่รองรับ
5	วัสดุ	Front glass, plastic body
6	SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)

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

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

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

```
result <- data.frame()

for (links in full_links[1:5]){
  ss_topic <- links %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

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

  tmp <- data.frame(
    attribute = ss_topic,
    value = ss_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
18	ชิปกราฟิก

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

	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
18	ชิปกราฟิก
19	หน่วยความจำ

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