

Mini Project 01 - IMDB web scraping

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

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

```
imdb <- read_html(url)
```

```
# movie title
titles <-imdb %>%
  html_nodes("h3.lister-item-header")%>%
  html_text2()
```

```
title[1:10]
```

ERROR: Error in title[1:10]: object of type 'closure' is not subsettable

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

```
# 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

MiniProject - Specphone web scraping

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

```
url <- read_html("https://specphone.com/OnePlus-11.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

```
# all one plus smartphone
oneplus_url <- read_html("https://specphone.com/brand/OnePlus")
```

```
links <- oneplus_url %>%
  html_nodes("li.mobile-brand-item a") %>% ## เว้นวรรคแล้ว a แปลว่า จงหา ตัว a ใน mobi:
  html_attr("href")
```

```
links
```

```
 '/OnePlus-Pad.html' · '/OnePlus-11R.html' · '/OnePlus-Nord-N20-5G.html' · '/OnePlus-10R.html' ·
```

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

```
full_links[1:10]
```

```
'https://specphone.com/OnePlus-Pad.html' · 'https://specphone.com/OnePlus-11R.html' ·
```

```

result <- data.frame()

for (link in full_links[1:10]){
  op_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()
  op_detail <- link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()
  tmp <- data.frame(attribute = op_topic,
                    value = op_detail)

  result <- bind_rows(result,tmp)
  print("Progress...")
}

```

```

[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."

```

```
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 ขีปประมวลผล
18 ขีปกราฟิก

```
write_csv(result, "result-op.csv")
```

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

	attribute	value
1	วันเปิดตัว	กุมภาพันธ์ 2566
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	258.00 x 189.00 x 6.50 มม.