

```
library (tidyverse)
library (rvest) ## Srape data from IoT
```

```
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_node จะเอาแค่ตัวแรกที่เจอ ถ้ามี s จะเอ  
  html_text2() ## Text เอยๆจะมี Character พิเศษด้วย
```

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

```
# Number of votes  
num_votes <- imdb %>%  
  html_node("p.sort-num_votes-visible") %>%  
  html_text2
```

```
# build a dataset  
df <- data.frame(  
  title = titles,  
  rating = ratings,  
  num_vote = num_votes  
)
```

Mini Project 2 - 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
✓ 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 <- read_html("https://specphone.com/Samsung-Galaxy-A04.html")
```

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

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

```
data.frame(attribute = att, value = detail)
```

A data.frame: 31 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ตุลาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	164.40 x 76.30 x 9.10 มม.
น้ำหนัก	192 กรัม
วัสดุ	Glass front, plastic back, plastic frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	-
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A
ประเภท	PLS LCD
ขนาดหน้าจอ	6.50 นิ้ว
ความละเอียด	720 x 1600 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Spreadtrum Unisoc SC9863A 1.6 GHz
ชิปกราฟิก	PowerVR GE8322
หน่วยความจำ	3 GB
ความจุ	32 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), AF ตัวที่ 2: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 5 MP, f/2.2
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	GLONASS, GALILEO, BDS
NFC	ไม่รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt

```
## All Samsung Smartphones
```

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

```
## Links to all Samsung Smartphones
```

```
links <-samsung_url %>%  
  html_nodes("li.mobile-brand-item a") %>% ## Spacebar -> จงหา a ที่อยู่ใน li.m  
  html_attr("href")
```

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

```
result <- data.frame()  
  
for (link in full_links){  
  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(attributes = ss_topic,  
                    value = ss_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 ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."  
[1] "Progress ..."
```

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

	attributes	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")
```