

# Web Scraping

## Mini Project 02 - Specphone Phone Database

### Load Related Library

```
library(tidyverse)

## — Attaching packages — tidyverse
1.3.2 —
## ✓ ggplot2 3.4.0      ✓ purrr 1.0.1
## ✓ tibble 3.1.8       ✓ dplyr 1.1.0
## ✓ tidyr 1.3.0        ✓ stringr 1.5.0
## ✓ readr 2.1.3        ✓ forcats 1.0.0
## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag() masks stats::lag()

library(rvest)

##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
## guess_encoding
```

### Read url

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

### Scape link for each model

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

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

### Build Dataset And Show Result

```
result <- data.frame()

for (link in full_links[1:5]){
  ss_topic <- link %>%
    read_html(link) %>%
```

```

    html_nodes("div.topic") %>%
    html_text2()

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

tmp <- data.frame(attribute = ss_topic,
                  value = ss_detail)

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

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

```

## Result

```

print(head(result))

##      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)

```