Mimi project 01 - IMDB web scraping

```
library(tidyverse)  #prep date/ prepare versualization / dplyr / ggplot2 is in he
library(rvest) #scrape data from the 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 code from website that we got the link
#ขอเข้าไปที่ link แล้วอ่านไฟล์ 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" widt .
```

```
#we want to find the title of the movies อยากดึงชื่อหนัง  Tiles
# Node is ตัวหนึ่งที่เราต้องการหา
# ถ้า Run ออกแปลว่าหาเจอ
imdb %>%
    html_node("h3.lister-item-header") #node ไม่มี s มันจะวิ่งเข้าไปหาอันแรกที่เจอ
```

```
{html_node}
<h3 class="lister-item-header">
[1] <span class="lister-item-index unbold text-primary">1.</span>
[2] <a href="/title/tt0111161/?ref_=adv_li_tt">The Shawshank Redemption</a>
[3] <span class="lister-item-year text-muted unbold">(1994)</span>
```

```
#จงดึงชื่อ text ออกมา
imdb %>%
    html_node("h3.lister-item-header") %>%
    html_text2() #use text2 มันจะลบพวกอัคระพิเศษออกออกไป ถ้าใช่ text1 ก็คือยังมีอยู่
```

'1. The Shawshank Redemption (1994)'

```
# NODES
# นี่คือขั้นตอนการดึง(scrape data) มาจากพวก wed site shows data ex.wiki pedia, imdb, med
imdb %>%
    html_nodes("h3.lister-item-header") %>%
    html_text2()
```

'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. Schindler\'s List (1993)' ·
'4. The Dark Knight (2008)' · '5. 12 Angry Men (1957)' · '6. The Godfather Part II (1974)' ·
'7. The Lord of the Rings: The Return of the King (2003)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·
'10. Fight Club (1999)' · '11. The Lord of the Rings: The Fellowship of the Ring (2001)' · '12. Forrest Gump (1994)' ·
'13. Il buono, il brutto, il cattivo (1966)' · '14. The Lord of the Rings: The Two Towers (2002)' · '15. GoodFellas (1990)' ·
'16. The Matrix (1999)' · '17. One Flew Over the Cuckoo\'s Nest (1975)' · '18. The Empire Strikes Back (1980)' ·
'19. Interstellar (2014)' · '20. Se7en (1995)' · '21. The Silence of the Lambs (1991)' · '22. Star Wars (1977)' ·
'23. The Green Mile (1999)' · '24. Saving Private Ryan (1998)' · '25. La vita è bella (1997)' ·
'26. Terminator 2: Judgment Day (1991)' · '27. Sen to Chihiro no kamikakushi (2001)' · '28. Cidade de Deus (2002)' ·
'29. It\'s a Wonderful Life (1946)' · '30. Shichinin no samurai (1954)' · '31. Seppuku (1962)' · '32. Whiplash (2014)' ·
'33. Gisaengchung (2019)' · '34. Back to the Future (1985)' · '35. Gladiator (2000)' · '36. The Prestige (2006)' ·
'37. The Departed (2006)' · '38. Apocalypse Now (1979)' · '39. Léon (1994)' · '40. Alien (1979)' ·
'41. The Usual Suspects (1995)' · '42. American History X (1998)' · '43. The Pianist (2002)' · '44. The Lion King (1994)' ·
'45. The Intouchables (2011)' · '46. Casablanca (1942)' · '47. Once Upon a Time in the West (1968)' ·
'48. Psycho (1960)' · '49. Hotaru no haka (1988)' · '50. Nuovo Cinema Paradiso (1988)'

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

titles[1:10] #show vector ข้อมูลที่เราเก็บมา
```

'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. Schindler\'s List (1993)' ·
'4. The Dark Knight (2008)' · '5. 12 Angry Men (1957)' · '6. The Godfather Part II (1974)' ·
'7. The Lord of the Rings: The Return of the King (2003)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·
'10. Fight Club (1999)'

```
#rating

#try with node
imdb %>%
    html_node("div.ratings-imdb-rating") %>%
    html_text2()
```

'9.3'

```
#rating uses nodes with "s"
imdb %>%
    html_nodes("div.ratings-imdb-rating") %>%
    html_text2() #This is character
```

'9.3' · '9.2' · '9.0' · '9.0' · '9.0' · '9.0' · '9.0' · '8.9' · '8.8' · '8.8' · '8.8' · '8.8' · '8.8' · '8.8' · '8.7' · '8.7' · '8.7' · '8.7' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5'

```
#transfer to numberic
#nodes
imdb %>%
    html_nodes("div.ratings-imdb-rating") %>%
    html_text2() %>% as.numeric
```

9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8 · 8.8 · 8.8 · 8.8 · 8.8 · 8.8 · 8.7 · 8.7 · 8.7 · 8.7 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.6 · 8.5

```
#ฝากค่าไว้ใน 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

```
# scrape vote ดึงโหวต
# number of vote start with one node
imdb %>%
    html_node("p.sort-num_votes-visible") %>%
    html_text2()
```

'Votes: 2,694,353 | Gross: $28.34M | Top 250: #1'

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

'Votes: 2,694,353 | Gross: $28.34M | Top 250: #1' · 'Votes: 1,869,755 | Gross: $134.97M | Top 250: #2' ·
'Votes: 1,362,665 | Gross: $96.90M | Top 250: #6' · 'Votes: 2,668,237 | Gross: $534.86M | Top 250: #3' ·
'Votes: 796,070 | Gross: $4.36M | Top 250: #5' · 'Votes: 1,278,369 | Gross: $57.30M | Top 250: #4' ·
'Votes: 1,856,151 | Gross: $377.85M | Top 250: #7' · 'Votes: 2,068,163 | Gross: $107.93M | Top 250: #8' ·
'Votes: 2,367,010 | Gross: $292.58M | Top 250: #14' · 'Votes: 2,139,750 | Gross: $37.03M | Top 250: #12' ·
'Votes: 1,885,626 | Gross: $315.54M | Top 250: #9' · 'Votes: 2,092,471 | Gross: $330.25M | Top 250: #11' ·
'Votes: 766,224 | Gross: $6.10M | Top 250: #10' · 'Votes: 1,676,057 | Gross: $342.55M | Top 250: #13' ·
'Votes: 1,169,082 | Gross: $46.84M | Top 250: #17' · 'Votes: 1,923,432 | Gross: $171.48M | Top 250: #16' ·
'Votes: 1,012,975 | Gross: $112.00M | Top 250: #18' · 'Votes: 1,299,626 | Gross: $290.48M | Top 250: #15' ·
'Votes: 1,850,095 | Gross: $188.02M | Top 250: #25' · 'Votes: 1,663,397 | Gross: $100.13M | Top 250: #19' ·
'Votes: 1,441,043 | Gross: $130.74M | Top 250: #22' · 'Votes: 1,372,052 | Gross: $322.74M | Top 250: #28' ·
'Votes: 1,309,985 | Gross: $136.80M | Top 250: #27' · 'Votes: 1,399,521 | Gross: $216.54M | Top 250: #24' ·
'Votes: 699,915 | Gross: $57.60M | Top 250: #26' · 'Votes: 1,105,754 | Gross: $204.84M | Top 250: #29' ·
'Votes: 770,402 | Gross: $10.06M | Top 250: #31' · 'Votes: 761,071 | Gross: $7.56M | Top 250: #23' ·
'Votes: 466,137 | Top 250: #21' · 'Votes: 348,505 | Gross: $0.27M | Top 250: #20' · 'Votes: 58,601 | Top 250: #45' ·
'Votes: 873,241 | Gross: $13.09M | Top 250: #42' · 'Votes: 816,175 | Gross: $53.37M | Top 250: #34' ·
'Votes: 1,214,053 | Gross: $210.61M | Top 250: #30' · 'Votes: 1,509,419 | Gross: $187.71M | Top 250: #37' ·
'Votes: 1,341,491 | Gross: $53.09M | Top 250: #41' · 'Votes: 1,332,990 | Gross: $132.38M | Top 250: #39' ·
'Votes: 672,329 | Gross: $83.47M | Top 250: #53' · 'Votes: 1,168,971 | Gross: $19.50M | Top 250: #35' ·
'Votes: 888,903 | Gross: $78.90M | Top 250: #51' · 'Votes: 1,090,969 | Gross: $23.34M | Top 250: #40' ·
'Votes: 1,128,546 | Gross: $6.72M | Top 250: #38' · 'Votes: 838,562 | Gross: $32.57M | Top 250: #32' ·
'Votes: 1,065,351 | Gross: $422.78M | Top 250: #36' · 'Votes: 865,193 | Gross: $13.18M | Top 250: #46' ·
'Votes: 575,607 | Gross: $1.02M | Top 250: #43' · 'Votes: 332,686 | Gross: $5.32M | Top 250: #48' ·
'Votes: 676,509 | Gross: $32.00M | Top 250: #33' · 'Votes: 280,844 | Top 250: #44' ·
'Votes: 263,928 | Gross: $11.99M | Top 250: #50'

```
#ฝากค่าไว้ที่ num_votes

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

```
#Right now we have 3 vector then we will combine 3 elements 1.rating 2.titles 3.v
#Build a dataset หัวใจของการ scraping data คือขั้นสุดท้ายเราก้จะได้ data set ก้อนนึ่งมาเพื่อส่งต่อใ
#ตัวอย่างก้คือ การเอาข้อมูลไปทำ charts, เอาไปหาข้อมูล inside บางอย่าง, หนังดราม่า 5 เรื่องแรกที่ทำเงินม

#Build data set.. Go home
df <- data.frame(
    title = titles,
    rating = ratings,
    num_vote = num_votes
)

#web scraping in R is easy if the website we put the data that was statics data,
head(df)
```

A data.frame: 6 × 3

	title	rating	num_vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,694,353 | Gross: $28.34M | Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,869,755 | Gross: $134.97M | Top 250: #2
3	3. Schindler's List (1993)	9.0	Votes: 1,362,665 | Gross: $96.90M | Top 250: #6
4	4. The Dark Knight (2008)	9.0	Votes: 2,668,237 | Gross: $534.86M | Top 250: #3
5	5. 12 Angry Men (1957)	9.0	Votes: 796,070 | Gross: $4.36M | Top 250: #5
6	6. The Godfather Part II (1974)	9.0	Votes: 1,278,369 | Gross: $57.30M | Top 250: #4

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

```
url = "https://specphone.com/Samsung-Galaxy-A04.html"
```

```
## Don't forgrt to read html
url = read_html("https://specphone.com/Samsung-Galaxy-A04.html")
```

```
#Topics
# the hardest part is to find which box(box model) that we need to pull out the h
url %>%
    html_nodes("div.topic") %>%
    html_text2()
```

'วันเปิดตัว' · 'วันวางจำหน่าย' · 'ขนาด' · 'น้ำหนัก' · 'วัสดุ' · 'SIM' · 'Technology' · '2G' · '3G' · '4G' · '5G' · 'ความเร็ว' · 'ประเภท' · 'ขนาดหน้าจอ' · 'ความละเอียด' · 'ระบบปฏิบัติการ' · 'ชิปประมวลผล' · 'ชิปกราฟิก' · 'หน่วยความจำ' · 'ความจุ' · 'Memory Card' · 'กล้องหลัก' · 'ความละเอียดวีดีโอ' · 'กล้องหน้า' · 'Bluetooth' · 'Wi-Fi' · 'USB' · 'GPS' · 'NFC' · 'ความจุ' · 'ประเภท'

```
#Details เราข้ามขั้นตอนนี้ได้ ไปทำข้างล่างรวมกันรอบเดียวได้เลย
url %>%
    html_nodes("div.detail") %>%
    html_text2()
```

'ตุลาคม 2565' · 'ยังไม่วางจำหน่าย' · '164.40 x 76.30 x 9.10 มม.' · '192 กรัม' · 'Glass front, plastic back, plastic frame' · 'รองรับ 2 ซิมการ์ด (nano sim, nano sim)' · 'HSPA 42.2/5.76 Mbps, LTE-A' · '850/900/1800/1900' · '850/900/1900/2100' · '850/900/1900/2100/2600' · '-' · 'HSPA 42.2/5.76 Mbps, LTE-A' · 'PLS LCD' · '6.50 นิ้ว' · '720 x 1600 pixels' · 'Android 12' · 'Spreadtrum Unisoc SC9863A 1.6 GHz' · 'PowerVR GE8322' · '3 GB' · '32 GB' · 'microSD (1)' · 'ตัวที่ 1: 50 MP, f/1.8, (wide), AF\nตัวที่ 2: 2 MP, f/2.4, (depth)' · '1080p@30fps' · 'ตัวที่ 1: 5 MP, f/2.2' · '5.0, A2DP, LE' · '802.11 a/b/g/n/ac, dual-b' · 'Type-C' · 'GLONASS, GALILEO, BDS' · 'ไม่รองรับ' · '5,000 mAh' · 'Non-removable Li-Po Batt'

```
#Build data frame
#ฝากค่าไว้ใน  att (attribute) & value  เพื่อสร้าง data  frame

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

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

```
#create data frame of Samsung galaxy A04
data.frame(attribute = att, value = value)
```

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)
ความละเอียดวีดีโอ	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

```
# Scrape all Samsung phones
#All Samsung smart phone
samsung_url <- read_html("https://specphone.com/brand/Samsung")
```

```
#1.link to all samsung smartphones
#2.in html code href is a contributon เราอยากจะดึง attribute ที่อยู่ใน li.mobile-brand-i
#3.type 1 space bar > a(child of li.mobile-brand-item)เพื่อเข้าไปหาลูกของ li.mobile-br
# when you find a > pull out data from href
samsung_url %>%
    html_node("li.mobile-brand-item a") %>%
    html_attr("href")
```

'/Samsung-Galaxy-S23-5G.html'

```
#ฝากค่าไว้ใน links // use web development knowledge
links <- samsung_url %>%
    html_node("li.mobile-brand-item a") %>%
    html_attr("href") #high per link refference
```

```
#nodes
links <- samsung_url %>%
    html_nodes("li.mobile-brand-item a") %>%
    html_attr("href")
# run links
links
```

'/Samsung-Galaxy-S23-5G.html' · '/Samsung-Galaxy-S23-Plus-5G.html' · '/Samsung-Galaxy-S23-Ultra-5G.html' · '/Samsung-Galaxy-M13.html' · '/Samsung-Galaxy-A23.html' · '/Samsung-Galaxy-A13.html' · '/Samsung-Galaxy-M32-5G.html' · '/Samsung-Galaxy-A12-Nacho.html' · '/Samsung-Galaxy-Pocket-Neo.html' · '/Samsung-Galaxy-Young.html' · '/Samsung-Galaxy-J1-Mini.html' · '/Samsung-Galaxy-A01-Core-1-16GB.html' · '/Samsung-Galaxy-A11.html' · '/Samsung-Galaxy-J2-Pro-2018.html' · '/Samsung-Galaxy-A12-2021.html' · '/Samsung-Galaxy-A21s-3-32GB.html' · '/Samsung-Galaxy-J5.html' · '/Samsung-Galaxy-J4.html' · '/Samsung-Galaxy-Core-2-Duos.html' · '/Samsung-Galaxy-Ace-Plus.html' · '/Samsung-Galaxy-A20.html' · '/Samsung-Galaxy-Chat.html' · '/Samsung-Galaxy-Gio.html' · '/Samsung-Galaxy-Tab-A7-Lite-LTE.html' · '/Samsung-Galaxy-Tab-A-10.5WIFI.html' · '/Samsung-Galaxy-Alpha.html' · '/Samsung-Galaxy-S3-Slim.html' · '/Samsung-Galaxy-S4-zoom.html' · '/Samsung-Galaxy-Xcover-2.html' · '/Samsung-Galaxy-Tab-8.9-3G-16GB.html' · '/Samsung-Galaxy-Tab-A8-LTE-2021.html' · '/Samsung-Galaxy-A8-2018.html' · '/Samsung-Galaxy-Tab4-8.0-wifi.html' · '/Samsung-Galaxy-M33-5G.html' · '/Samsung-Galaxy-A50.html' · '/Samsung-Galaxy-E7.html' · '/Samsung-Galaxy-S6.html' · '/Samsung-Galaxy-S20-FE.html' · '/Samsung-Galaxy-Tab-S4-WIFI.html' · '/Samsung-Galaxy-S7.html' · '/Samsung-Galaxy-Note-5-Exynos.html' · '/Samsung-Galaxy-TabPRO-12.2-LTE.html' · '/Samsung-Galaxy-S4-Active.html' · '/Samsung-Galaxy-Tab-Active-3.html' · '/Samsung-Galaxy-Tab-S3-9.7.html' · '/Samsung-Galaxy-S6-edge.html' · '/Samsung-Galaxy-Note-4-Exynos.html' · '/Samsung-Galaxy-Round.html' · '/Samsung-Galaxy-Note-20-Ultra-5G.html' · '/Samsung-ATIV-Q.html' · '/Samsung-ATIV-Smart-PC-PRO.html' · '/Samsung-Galaxy-S22-Ultra12-128GB.html' · '/Samsung-Galaxy-Z-Flip-5G.html' · '/Samsung-Galaxy-Z-Flip.html' · '/Samsung-Galaxy-Tab-S8-Ultra-5G.html' · '/Samsung-Galaxy-S21-Ultra-16-512GB.html' · '/Samsung-Galaxy-S10-Plus-Ram-12GB.html' · '/Samsung-Galaxy-Z-Fold-3.html' · '/Samsung-Galaxy-Z-Fold4.html' · '/Samsung-Galaxy-Z-Fold-2-5G.html'

```
#Read html2
url2 = "https://specphone.com"

print(url2)
```

```
[1] "https://specphone.com"
```

```
full_links <- read_html(url2)
```

```
print(url2)
```

```
[1] "https://specphone.com"
```

```
#we want to click on the link use str_c(..) or paste0("https://specphone.com") เพื่

paste0("https://specphone.com", links[1:5])
```

'https://specphone.com/Samsung-Galaxy-S23-5G.html' ·
'https://specphone.com/Samsung-Galaxy-S23-Plus-5G.html' ·
'https://specphone.com/Samsung-Galaxy-S23-Ultra-5G.html' ·
'https://specphone.com/Samsung-Galaxy-M13.html' · 'https://specphone.com/Samsung-Galaxy-A23.html'

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

```
full_links
```

'https://specphone.com/Samsung-Galaxy-S23-5G.html' ·
'https://specphone.com/Samsung-Galaxy-S23-Plus-5G.html' ·
'https://specphone.com/Samsung-Galaxy-S23-Ultra-5G.html' ·
'https://specphone.com/Samsung-Galaxy-M13.html' · 'https://specphone.com/Samsung-Galaxy-A23.html' ·
'https://specphone.com/Samsung-Galaxy-A13.html' · 'https://specphone.com/Samsung-Galaxy-M32-5G.html' ·
'https://specphone.com/Samsung-Galaxy-A12-Nacho.html' ·
'https://specphone.com/Samsung-Galaxy-Pocket-Neo.html' ·
'https://specphone.com/Samsung-Galaxy-Young.html' · 'https://specphone.com/Samsung-Galaxy-J1-Mini.html' ·
'https://specphone.com/Samsung-Galaxy-A01-Core-1-16GB.html' ·
'https://specphone.com/Samsung-Galaxy-A11.html' · 'https://specphone.com/Samsung-Galaxy-J2-Pro-2018.html' ·
'https://specphone.com/Samsung-Galaxy-A12-2021.html' ·
'https://specphone.com/Samsung-Galaxy-A21s-3-32GB.html' · 'https://specphone.com/Samsung-Galaxy-J5.html' ·
'https://specphone.com/Samsung-Galaxy-J4.html' · 'https://specphone.com/Samsung-Galaxy-Core-2-Duos.html' ·
'https://specphone.com/Samsung-Galaxy-Ace-Plus.html' · 'https://specphone.com/Samsung-Galaxy-A20.html' ·
'https://specphone.com/Samsung-Galaxy-Chat.html' · 'https://specphone.com/Samsung-Galaxy-Gio.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-A-10.5WIFI.html' ·
'https://specphone.com/Samsung-Galaxy-Alpha.html' · 'https://specphone.com/Samsung-Galaxy-S3-Slim.html' ·
'https://specphone.com/Samsung-Galaxy-S4-zoom.html' ·
'https://specphone.com/Samsung-Galaxy-Xcover-2.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-A8-LTE-2021.html' ·
'https://specphone.com/Samsung-Galaxy-A8-2018.html' ·
'https://specphone.com/Samsung-Galaxy-Tab4-8.0-wifi.html' ·
'https://specphone.com/Samsung-Galaxy-M33-5G.html' · 'https://specphone.com/Samsung-Galaxy-A50.html' ·
'https://specphone.com/Samsung-Galaxy-E7.html' · 'https://specphone.com/Samsung-Galaxy-S6.html' ·
'https://specphone.com/Samsung-Galaxy-S20-FE.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-S4-WIFI.html' · 'https://specphone.com/Samsung-Galaxy-S7.html' ·
'https://specphone.com/Samsung-Galaxy-Note-5-Exynos.html' ·
'https://specphone.com/Samsung-Galaxy-TabPRO-12.2-LTE.html' ·
'https://specphone.com/Samsung-Galaxy-S4-Active.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-Active-3.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-S3-9.7.html' ·
'https://specphone.com/Samsung-Galaxy-S6-edge.html' ·
'https://specphone.com/Samsung-Galaxy-Note-4-Exynos.html' ·
'https://specphone.com/Samsung-Galaxy-Round.html' ·
'https://specphone.com/Samsung-Galaxy-Note-20-Ultra-5G.html' · 'https://specphone.com/Samsung-ATIV-Q.html' ·
'https://specphone.com/Samsung-ATIV-Smart-PC-PRO.html' ·
'https://specphone.com/Samsung-Galaxy-S22-Ultra12-128GB.html' ·
'https://specphone.com/Samsung-Galaxy-Z-Flip-5G.html' · 'https://specphone.com/Samsung-Galaxy-Z-Flip.html' ·
'https://specphone.com/Samsung-Galaxy-Tab-S8-Ultra-5G.html' ·
'https://specphone.com/Samsung-Galaxy-S21-Ultra-16-512GB.html' ·
'https://specphone.com/Samsung-Galaxy-S10-Plus-Ram-12GB.html' ·
'https://specphone.com/Samsung-Galaxy-Z-Fold-3.html' · 'https://specphone.com/Samsung-Galaxy-Z-Fold4.html' ·
'https://specphone.com/Samsung-Galaxy-Z-Fold-2-5G.html'

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

```
#For loop try 10 links
# Don't forget to read_html
result <- data.frame()

    for (link in full_links[1:10]) {
        ss_topic <- link %>%  #เราจะวิ่งเข้าไปหา div.topic
        read_html %>%
        html_nodes("div.topic") %>%
        html_text2()

    ss_detail <- link %>%
        read_html %>%
        html_nodes("div.detail") %>%
        html_text2()
    tmp <- data.frame(attribute = ss_topic, value = ss_detail)

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

print(result)
```

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

```
#write CSV
write_csv(result, "result_ss_phone.csv")
```

```
print(head(result),5)
```

```
    attribute
1     วันเปิดตัว
2 วันวางจำหน่าย
3         ขนาด
4       น้ำหนัก
5         วัสดุ
6          SIM

1
2                                                                          ย
3                                                                 146.30 x 70.9
4
5 Glass front (Gorilla Glass Victus 2), glass back (Gorilla Glass Victus 2), al
6                                                    รองรับ 2 ซิมการ์ด (nano sim
```