

[웹 데이터 수집]

1. 정적 웹 크롤링: 단일 페이지 크롤링

- 필요한 패키지 설치 및 로드: `install.packages("rvest") / library(rvest)`

2. 동적 웹 크롤링: 스크롤 다운, 로그인 이후 화면, 버튼...

(1) 사전 준비

- 필요한 파일들 사이트에서 다운로드하여 한 폴더에 모아두기:

`geckodriver.exe / chromedriver.exe / selenium-server-standalone-3.141.59.jar`

- 필요한 패키지 설치 및 로드:

`install.packages("RSelenium") / library(httr) / library(rvest) / library(RSelenium)`

- 다운로드한 파일들을 모아둔 폴더에서 cmd창을 열어서 아래와 같이 입력

`'java -Dwebdriver.gecko.driver="geckodriver.exe" -jar selenium-server-standalone-3.141.59.jar -port 4445'`

(2) remoteDriver

```
ex) remDr <- remoteDriver(port=4445L,           # 포트번호
                           browserName = 'chrome') # 사용할 브라우저
remDr$open()                                   # 실행 결과를 실시간으로 확인할 새 브라우저 창 열기
remDr$navigate("http://www.youtube.com")       # 새 브라우저 창에 www.youtube.com 사이트 열기
```