########################정적데이터#######################

R : Crawling & Scrapping

-Naver 영화 댓글

: 클라이언트가 request를 보내고 서버 프로그램이 요청을 받아,

서버 프로그램이 결과 HTML page를 생성해서 클라이언트에게 response로 전달한다 => 클라이언트는 Selector와 xpath를 이용해서 전달된 HTML내에 있는 필요한데이터를 추출한다

-로튼 토마토 : 위오 ㅏ 같음

- kakao api(이미지)=OPEN API

: KAKAO 이미지 검색 API(Open API를 이용해서 데이터를 구축하는 방식)

=> 둘 다 방식이 같음

클라이언트(R)가 요청함 > 서버프로그램이 요청받으면 결과 JSON 문자열을 생성 > 이 결과 데이터를 data frame으로 받는다 > data frame으로 결과를 추출한다

############# 이 두가지 방식으로 해결할 수 없는 상황 ############

Open API를 사용하지 않는(공개되지 않는 API를 사용) AJAX 방식의 Web Page는 위의 방법으로 데이터를 가져올 수 없다

ex) 도서검색

1차적으로 웹페이지에 아무것도 없고, 사용자가 임의의 조작을 해야 페이지에서 데이터를 가져올 수 있다.

= 동적 데이터 => Selenium을 이용한 동적 데이터 Crawling

-Selenium server program 을 다운로드한다

-Selenium이 Chrome browser를 자동으로 제어하게 된다

-Chrome을 자동으로 제어하려면 chrome driver가 필요하다

-이 chrome driver 프로그램을 selenium이 사용할 수 있도록 PATH 환경변수에 경로를 설정

-Selenium 서버를 기동하자

java -jar 파일명 - port 4445