해외에서 흥행한

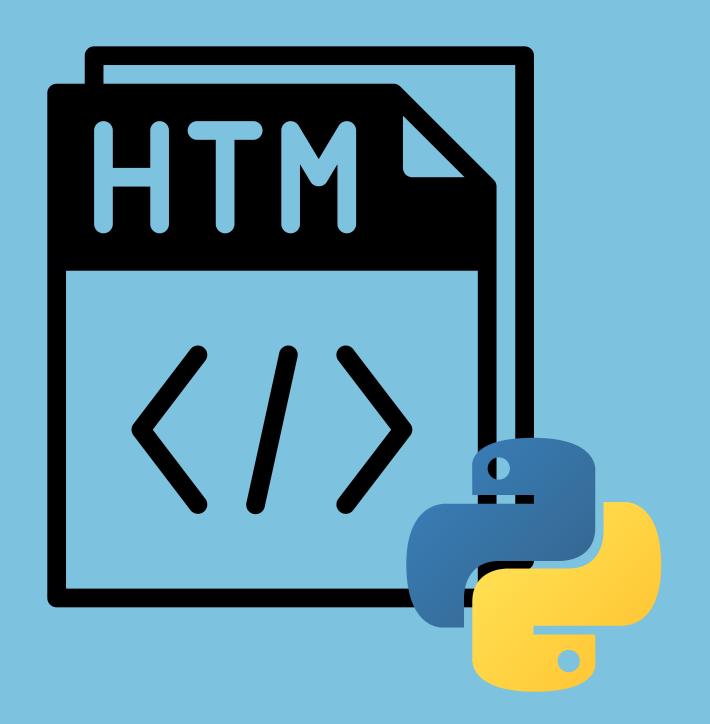
### K-Content 흥행 요인 분석

https://beatdrum.netlify.app/index.html





#### 북치고 장고 치우고





주제 선정 이유 및 방향

리뷰 크롤링

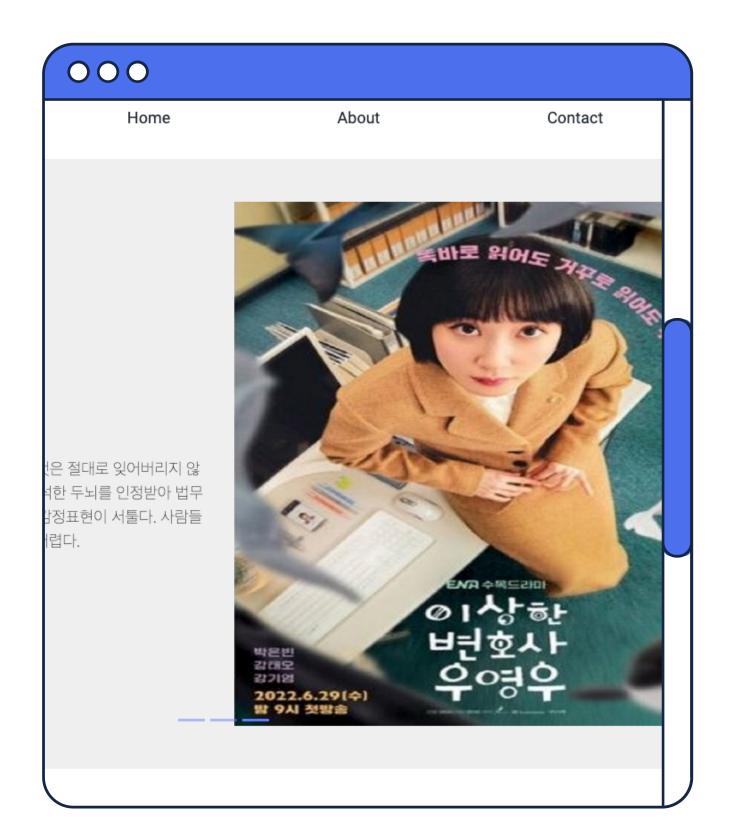
데이터 전처리 및 기술 사용

웹페OI지 구현

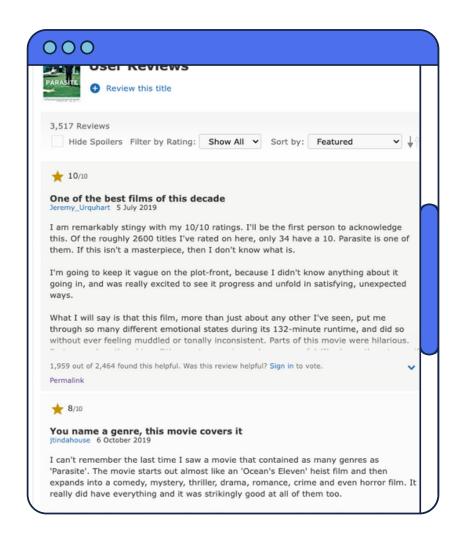
느낌점

### 주제 선정 이유 및 방향

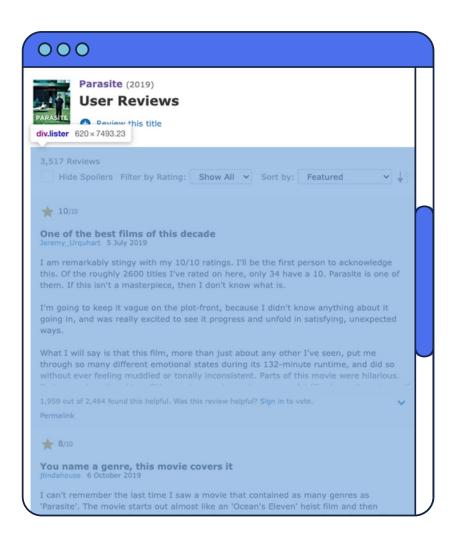
- 1 코로나-19 발생 후 미디어 매체의 이용량 증가
- 2 K-컨텐츠의 해외 접속량 증가
- 한국정서와 다른 외국인들의 리뷰를 바탕으로 흥행요인 분석



# 리뷰 크롤링



IMDB.com 리뷰사이트 접속 분석 할 컨텐츠(6가지)의 리뷰 페이지에 접속



리뷰 부분 크롤링 웹 페이지 검사 도구로 제목, 리뷰, 및 더보기 버튼의 XPATH를 탐색

```
000
from selenium.webdriver.firefox.firefox_binary import FirefoxBinary
import time
from urllib.parse import urljoin
import pandas as pd
import requests
import os
# Mac OS의 경우와 그 외 OS의 경우로 나누어 설정
if os.name == "posix":
   driver = FirefoxBinary("/Applications/Firefox.app/Contents/MacOS/fi/
    browser = webdriver.Firefox(firefox_binary=driver)
   driver = FirefoxBinary("C:/Program Files/Mozilla Firefox/firefox.ex
    browser = webdriver.Firefox(firefox_binary=driver)
movie_list = [
    "tt6751668",
    "tt10919420",
    "tt10633456",
    "tt5700672",
    "tt20869502",
    "tt14518756",
review_list = []
rating_list = []
# movie_list 사용하여 리뷰 크롤링 시작, 리뷰 평점 높은 순서대로 크롤링
for movie in movie_list:
    url = (
        "https://www.imdb.com/title/"
```

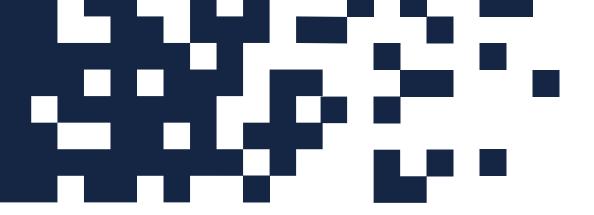
imdb\_review\_movies.py Selenium, BeautifulSoup, requests 등을 사용하여 크롤링

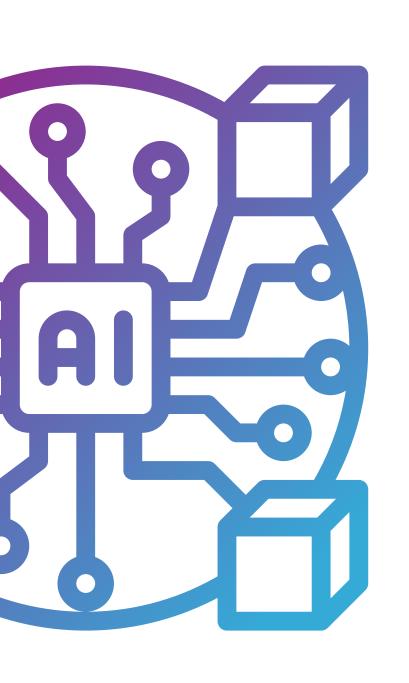
### 데이터 전처리



- 1 모든 리뷰를 하나의 문자열로 합친다.
- Adtext\_to\_word\_sequence()를 이용해 단어 단위로 분리
- 물용어 제거 (영어가 아닌 단어, 길이가 4이하인 단어)
- 4 표제어 추출
- 3 품사 테그 후 분리

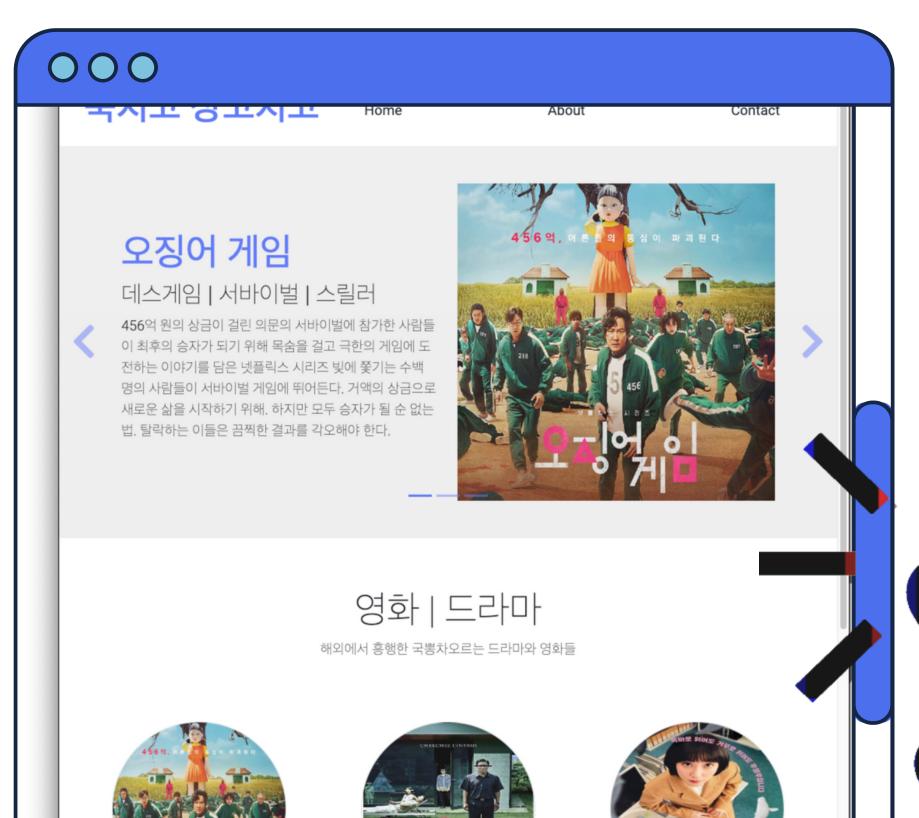
#### 기술 사용

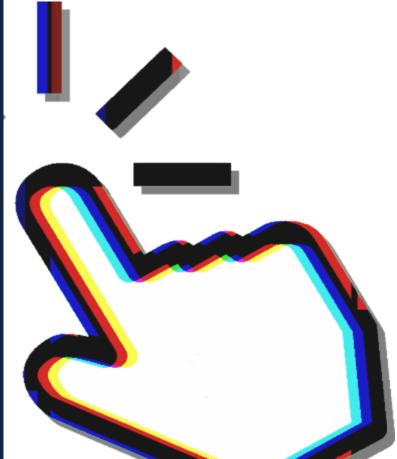




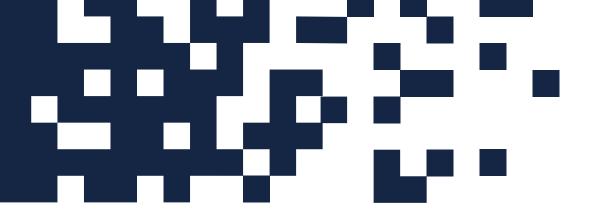
- 1 GloVe를 통하여 모든 단어 벡터화
- **2** 평균벡터를 연관성이 높은 단어벡터로 판단 및 추출
- 평균벡터와 각 단어 벡터 간의 코사인 유사도 비교
- 4 코사인 유사도가 높은 순으로 단어를 정렬
- 5 품사가 명사인 단어만 추출 및 저장

## 웹메이지 구현 1





할 수 있어 뜻깊었습니다!



유명상 박정현 새로운것을 배워가며 프로젝트를 완성

정현님, 저 너무 힘들어요.