

개발 과정

1. 데이터 크롤링

"크롤링"이란?

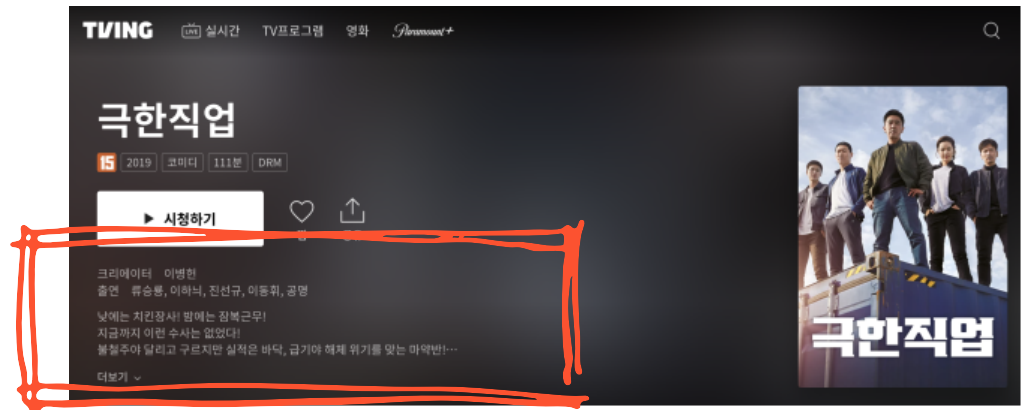
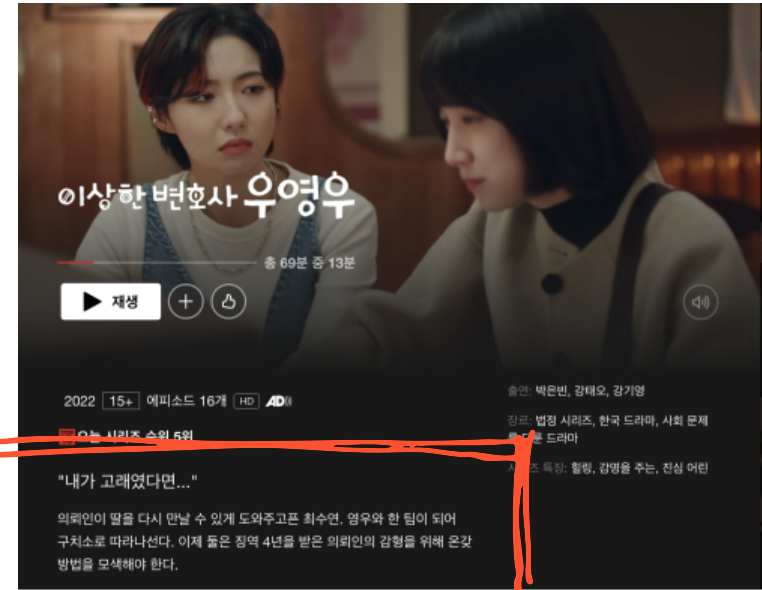
웹 페이지에서 데이터를 추출해 내는 과정을 말합니다.

국내 ott 사이트인 "넷플릭스"와 "티빙"에서 데이터 크롤링 하였습니다.

크롤링으로 가져온 데이터

1. 드라마 시놉시스
2. 이미지
3. 드라마 제목

beautifulsoup과 Selenium을 사용

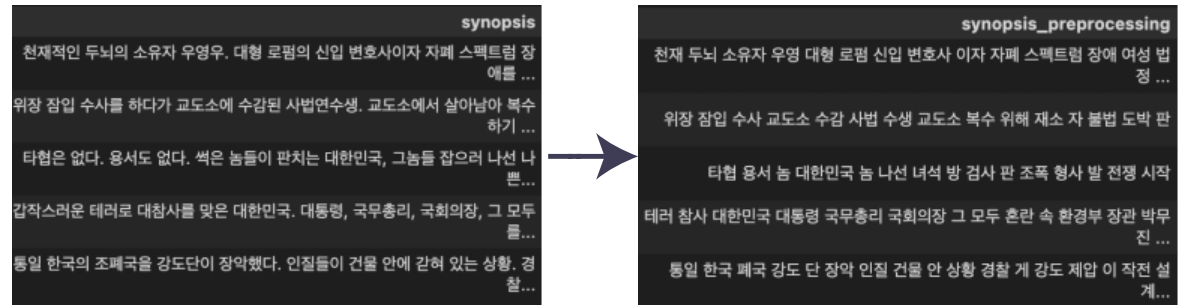


2. 모델 개발

크롤링을 하여 가져온 정보들 중
"드라마 시놉시스" 정보를 이용하여 드라마
 추천 시스템 머신러닝을 개발하였습니다.

1. 한국어 텍스트를 처리

한국어 텍스트를 처리하기 위해 **"KoNLPy"** 라이브러리
 를 이용하여 문장을 **명사 단위**로 분리시켰습니다.



2. Feature 벡터화

문서에서 추출된 단어들을 모으로, 인덱스를 매깁니다. 이후 단어들이 나
 타난 횟수를 기준으로 **Feature Vector**를 생성합니다.

3. 코사인 유사도

코사인 유사도를 이용해서 영화간의 비슷한 정도를 구합니다.
 이후 유사도가 높은 순서대로 정렬하여 웹으로 넘겨줍니다.
 (추천 정확도가 그닥 높지는 않은 것 같네요..)

```
{
  '리갈하이': [1, 1, 0, 0, 0, 1],
  '손 the guest': [1, 1, 0, 0, 1, 1],
  '우아한 가': [1, 1, 0, 1, 0, 0],
  '하이에나': [1, 1, 0, 1, 1, 0],
  '무법변호사': [1, 1, 0, 0, 1, 0],
  '이상한 변호사 우영우': [1, 0, 0, 0, 1, 0],
  '내일 지구가 망해버렸으면 좋겠어': [0, 0, 0, 0, 1, 0],
  '종이의 집 공동경제구역': [0, 0, 0, 0, 1, 0],
  '경여년': [1, 1, 0, 0, 1, 1],
  '바벨': [1, 1, 0, 1, 0, 0]}

```

3. 웹 개발

인터페이스로는 웹을 이용하였습니다.

추천된 영화를 보기 가장 좋은 ott 별 가성비를
사용자에게 제공하는 기능을 두어 다른
추천시스템들과는 차별을 두었습니다.

사용자가 드라마를 입력하면 가장 그 드라마와
가장 유사한 드라마 9개를 추천해 줍니다.



OTT 별 한개의 드라마를 시청하는데 들어가는 비용

Seezn - 1500원 / 1 drama
Watcha - 7900원 / 1 drama
Wavve - 7900원 / 1 drama
Netflix - 1357원 / 1 drama
Tving - 1580원 / 1 drama

추천된 영화를 모두 보기에 가성비가 가장 좋은 OTT는 "Netflix"입니다.

