

찾아 줘요!

영화 제목

START

영화 🐾
보고싶
다옹~

20231005 영화몰라요 박소원 배진호 오승준 윤성원



Q

어떤 영화 봤나요?

장난감들이 살아
움직이면서 모험하는 영화



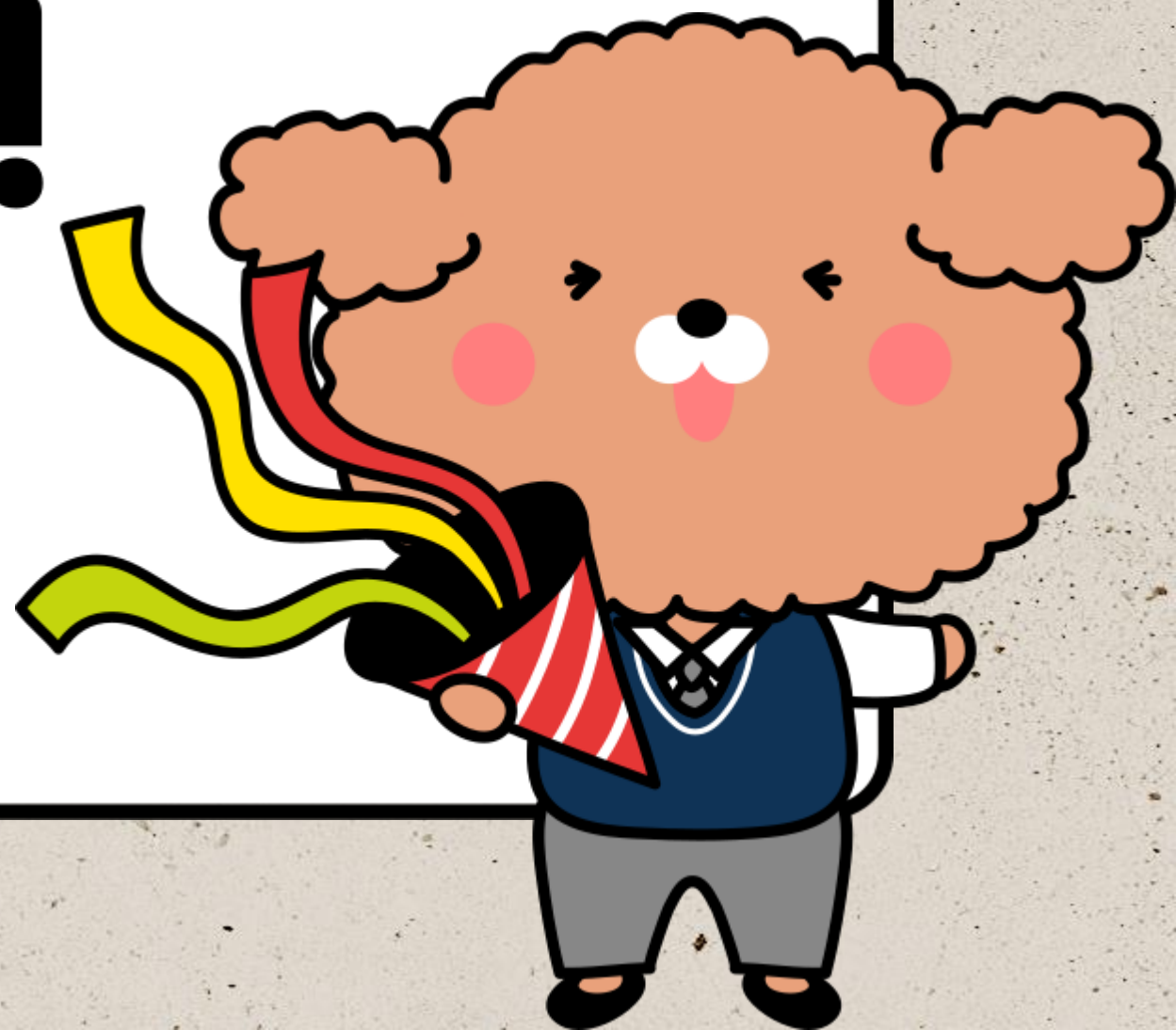
기억이
나듯말듯

A

강이가 예상한
영화제목은.. 바로!

토이스토리!

멍아!



Q

어떤 영화였나요?

인공지능이랑 사랑에
빠져서 상처를 회복
하는 영화!

좀 어려워



A

야옹이가 예상한
영화 제목은 바로!

A.I. (Artificial Intelligence)

다 귀엽아!
영화야!



그러나
정답은...

그녀(her)





목차

- 주제 선정 이유
- 크롤링
- 감정분석
- kkma & cosine
- 유사도 분석 & 서비스 구현
- 역할분담



NEW

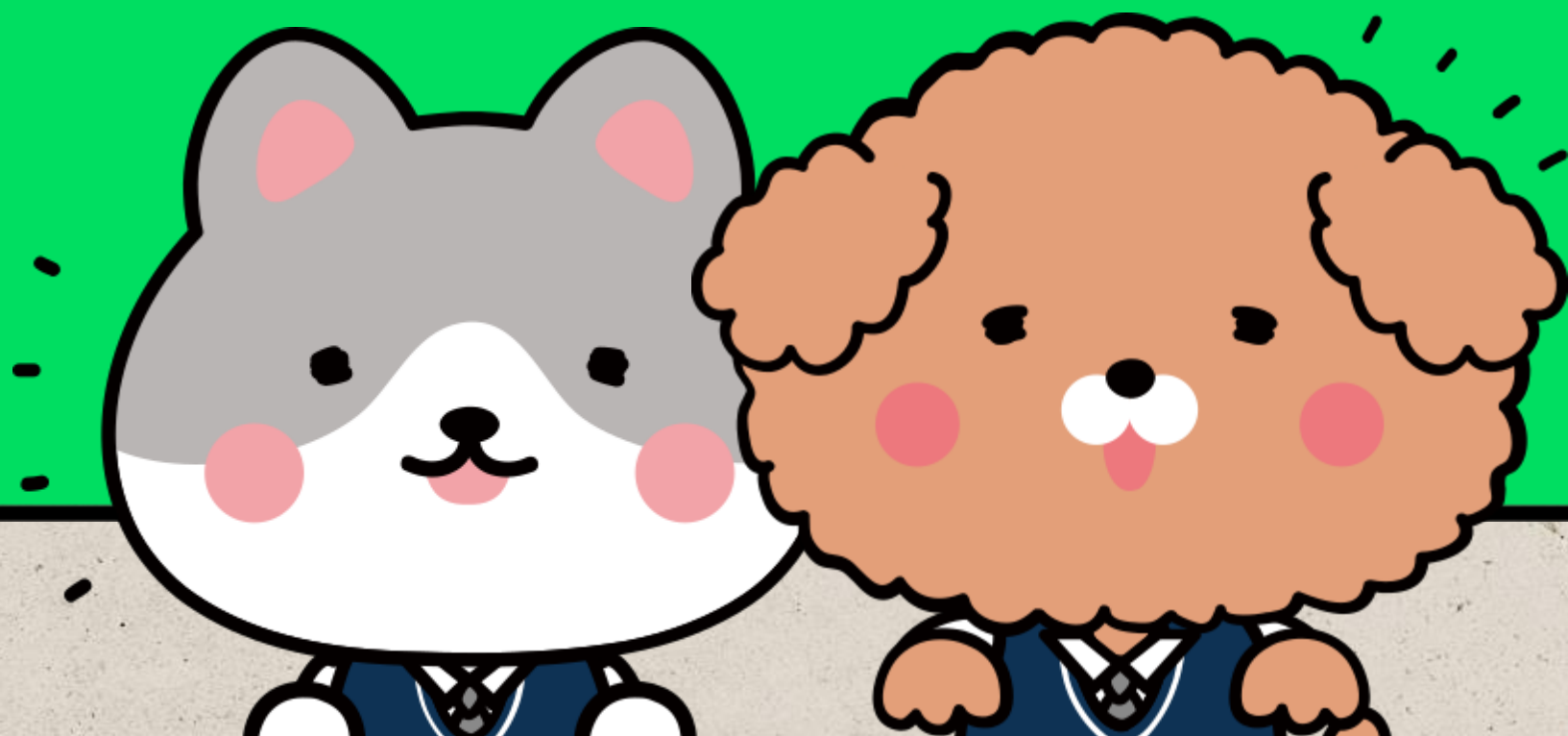
영화 제목이 기억 안날 때
줄거리를 적으면
제목도 알려주는
프로그램

다 계획이

있구나



크롤링



how

네이버 시리즈 온 - 인기순

NAVER 시리즈온

웹소설 · 만화 > 로그인

영화

방송

랭킹

무료

이벤트

멤버십

첫 달 무료

MY 카테고리 ▾

작품, 배우, 감독명을 입력하세요.



전체

액션

코미디

드라마

멜로

공포/스릴러

SF/판타지

애니메이션

다큐멘터리

독립영화

공연실황

HOT & NEW

무료

할인

소장

가치봄

영어자막

디즈니

워너브라더스

소니픽쳐스

NBC유니버설

20세기폭스

파라마운트

전체(28,343)

업데이트순

평점순

인기순

가격별 보기 ▾

조건별 보기 ▾

```
driver = webdriver.Chrome(service=service,options=options)
driver.get("https://serieson.naver.com/v3/movie/products?tag=&sortType=POPULARITY_DESC&price=all")
driver.implicitly_wait(10)
time.sleep(3)
```


what

제목, 평점, 감독, 배우, 개봉일자, 줄거리 크롤링



범죄도시3

★ 7.6 | 감독 이상용 | 주연 마동석, 이준혁, 아오키 무네타카, 이범수,... | 2023.05.31.

대체불가 괴물형사 마석도, 서울 광수대로 발탁! 베트남 납치 살해범 검거 후 7년 뒤, '마석도'(마동석)는 새로운 팀원들과 함께 살인사건을 조사한다. 사건 조사 중, '마석도'는 신종 마약 사건이 연루되었음을 알게 되고 수사를 확대한다. 한편, 마약 사건의 배후인 '주성철'(이준혁)은 계속해서 판을 키워가고 약을 유통하던 일본 ...

구매 11,000©



엘리멘탈(패키지: 자막판+더빙판+부가영상+코멘터리)

★ 9.4 | 감독 피터 손 | 주연 레아 루이스, 마무두 아티 | 2023.06.14.

불, 물, 공기, 흙 4개의 원소들이 살고 있는 '엘리멘트 시티' 재치 있고 불처럼 열정 넘치는 '앰버'는 어느 날 우연히 유쾌하고 감성적이며 물 흐르듯 사는 '웨이드'를 만나 특별한 우정을 쌓으며, 지금까지 믿어온 모든 것들이 흔들리는 새로운 경험을 하게 되는데... 웰컴 투 '엘리멘트 시티'!

구매 10,900©

how

크롤링 횟수 정하기

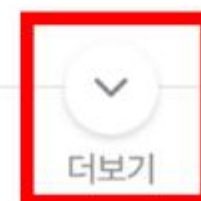


앤트맨과 와스프: 퀀텀매니아(세트: 1~3편)

★ 7.0 | 감독 페이튼 리드 | 주연 폴 러드, 에반젤린 릴리, 조나단 메이저스,... | 2023.02.15.

슈퍼히어로 파트너인 '스캇 랭'(폴 러드)과 '호프 반 다인'(에반젤린 릴리), 호프의 부모 '재닛 반 다인'(미셸 파이퍼)과 '헝크 핼'(마이클 더글라스), 그리고 스캇의 딸 '캐시 랭'(캐서린 뉴튼)까지 미지의 '양자 영역' 세계 속에 빠져버린 '앤트맨 패밀리'. 그 곳에서 새로운 존재들과 무한한 우주를 다스리는 정복자 '칸'을 만나며, 그 ...

구매 14,850©



더보기 버튼을 눌러도 맨 위 데이터가 사라지지 않음

→ 스크롤 후 크롤링 시도

→ 메모리 사용량이 커지며 느려져

220회 [더보기] 버튼 클릭 후 크롤링 진행

how

크롤링 이슈 해결



범죄도시3

★ 7.6 | 감독 이상용 | 주연 마동석, 이준혁, 아오키 무네타카, 이범수, ... | 2023.05.31.

대체불가 괴물형사 마석도, 서울 광수대로 발탁! 베트남 납치 살해범 검거 후 7년 뒤, '마석도'(마동석)는 새로운 팀원들과 함께 살인사건을 조사한다. 사건 조사 중, '마석도'는 신종 마약 사건이 연루되었음을 알게 되고 수사를 확대한다. 한편, 마약 사건의 배후인 '주성철'(이준혁)은 계속해서 판을 키워가고 **약을 유통하던 일본 ...**

구매 11,000₩

▼<p class="ExpandedPoster_synopsis__KyuZ7">

"대체불가 괴물형사 마석도, 서울 광수대로 발탁! 베트남 납치 살해범 검거 후 7년 뒤, '마석도'(마동석)는 새로운 팀원들과 함께 살인사건을 조사한다. 사건 조사 중, '마석도'는 신종 마약 사건이 연루되었음을 알게 되고 수사를 확대한다. 한편, 마약 사건의 배후인 '주성철'(이준혁)은 계속해서 판을 키워가고 **약을 유통하던 일본 조직과 '리키'(아오키 무네타카)까지 한국에 들어오며 사건의 규모는 점점 더 커져가는데... 나쁜 놈들 잡는 데 이유 없고 제한 없다. 커진 판도 시원하게 싹 쓸어버린다!"**

**메인 화면에서 줄거리가 길면
잘려서 보이지 않는 현상**

- 개발자 도구에서 전체 줄거리 확인
- 메인 페이지에서 크롤링

how

크롤링 중복 처리

'범죄도시'에 대한 검색결과입니다.



범죄도시2

영화 | 2022 | 마동석, 손석구, 최귀화, 박지환, 허동원, 하준, ...

가리봉동 소탕작전 후 4년 뒤, 금천서 강력반은 베트남으로 도주한 용의자를 인!
현지 용의자에게서 수상함을 느끼고, 그의 뒤에 무자비한 악행을 벌이는 '강해상

구매 6,050©



범죄도시2

영화 | 2022 | 마동석, 손석구, 최귀화, 박지환, 허동원, 하준, ...

가리봉동 소탕작전 후 4년 뒤, 금천서 강력반은 베트남으로 도주한 용의자를 인!
현지 용의자에게서 수상함을 느끼고, 그의 뒤에 무자비한 악행을 벌이는 '강해상

대여 2,750©

중복된 영화가 있음

→ 영화제목 : [평점, 감독, 배우, 개봉일자, 줄거리] 형태의
dict type으로 저장하여 영화 제목의 중복이 없도록 함

```
total={}

main=driver.find_elements(By.CSS_SELECTOR, ".ExpandedListCollection_item___Swc_")

for i in tqdm(main):

    result = i.get_attribute("innerHTML")
    from bs4 import BeautifulSoup
    bs = BeautifulSoup(result,"html.parser")

    try:
        name = bs.select_one('.ExpandedPoster_title__P7ff9').text
        star = bs.select_one('.ExpandedPoster_info_item__ZsGjd').text.split('\n')[-1]
        director = bs.select_one('.ExpandedPoster_info_item__ZsGjd.ExpandedPoster_directors__JBSE1').text
        actor = bs.select_one('.ExpandedPoster_info_item__ZsGjd.ExpandedPoster_actors__1lZnU').text
        date_ = bs.select_one('.ExpandedPoster_info_item__ZsGjd')[-1].text
        content=bs.select_one('.ExpandedPoster_synopsis__KyuZ7').text

        total[name]=[star,director,actor,date_,content]

    except:
        pass
```


how

크롤링 전처리 과정

```
result = [[k,v[0],v[1],v[2],v[3],v[4]] for k,v in total.items()]

result=pd.DataFrame(result,columns=['제목','평점','감독','배우','개봉일자','줄거리'])

result = result.applymap(lambda x : x.replace(",",""))
result = result.applymap(lambda x : x.replace("\n",""))
result = result.applymap(lambda x : x.replace("\r",""))
result['감독']=result['감독'].apply(lambda x : x.replace("감독",""))
result['평점']=result['평점'].apply(lambda x : x.replace("별점",""))
result['배우']=result['배우'].apply(lambda x : x.replace("주연",""))
```

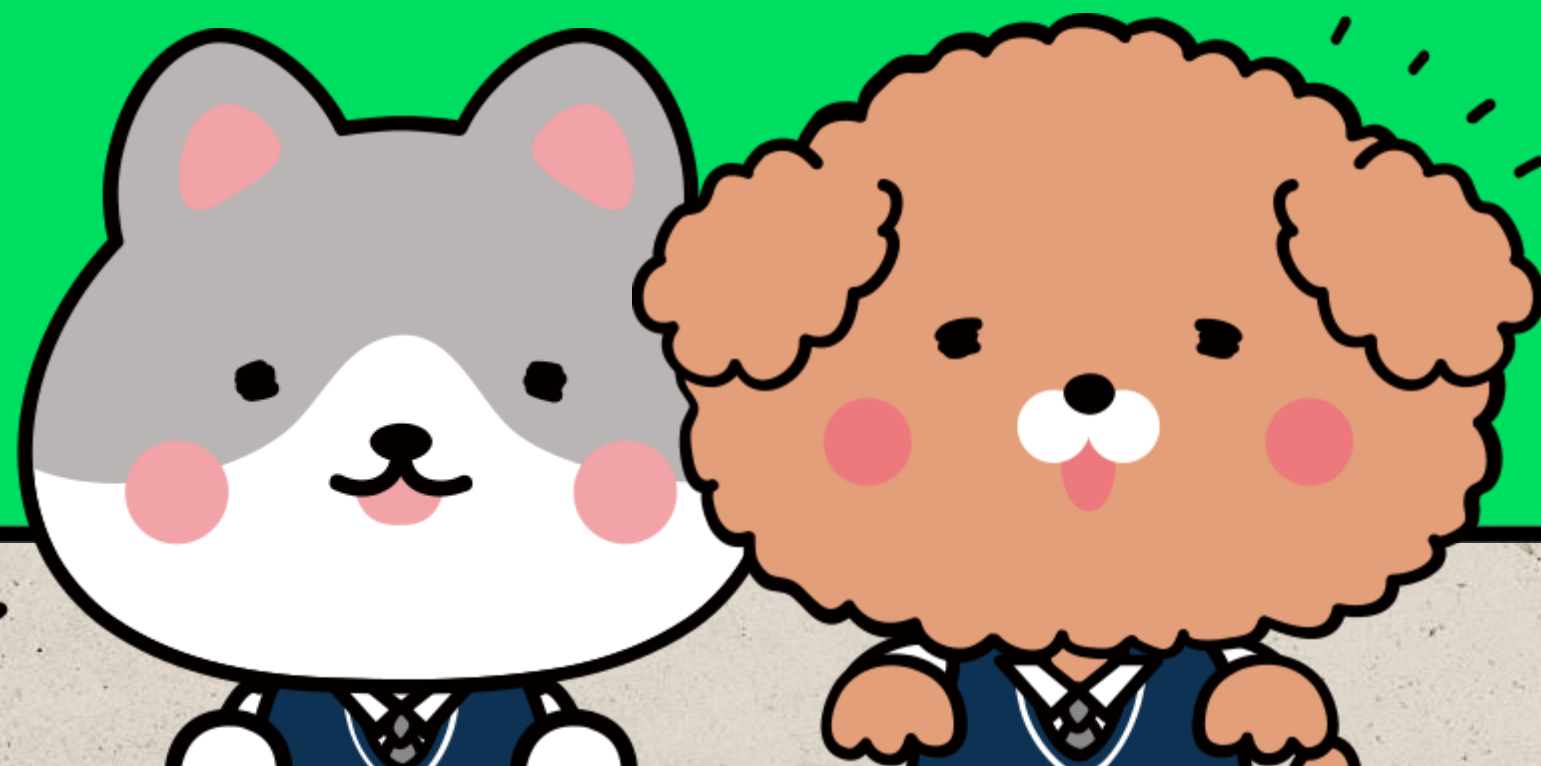
```
result.to_csv("movie.csv",encoding='utf-8',index=False)
```

범죄도시3

★ 7.6 | 감독 이상용 | 주연 마동석, 이준혁, 아오키 무네타카, 이범수,...

- 데이터프레임 형태로 변환
- 전처리
 - “,” , “\n” , “\r” 삭제
 - 각 항목이름 삭제
- 최종 5000여개 행
5196 rows × 6 columns
- csv 파일로 저장하여 분석 진행

감정분석



how

감정분석

	제목	평점	감독	배우	개봉일자	줄거리
0	범죄도시3	7.6	이상용	마동석 이준혁 아오키 무네타카 이범수 김민재 이지훈 김도건 고규필 전석호 안세호	2023.05.31.	대체불가 괴물형사 마석도 서울 광수대로 발탁! 베트남 납치 살해범 검거 후 7년 뒤...
1	엘리멘탈(패키지: 자막판+더빙판+부가영상+코멘터리)	9.4	피터 손	레아 루이스 마무두 아티	2023.06.14.	불 물 공기 흙 4개의 원소들이 살고 있는 '엘리멘트 시티' 재치 있고 볼거리 열정...
2	공조2: 인터내셔널	8.3	이석훈	현빈 유해진 윤아 다니엘 헤니 진선규 장영남 박훈 임성재 윤상화 박민하	2022.09.07.	공조 이즈 백! 이번엔 삼각 공조다! 남한으로 숨어든 글로벌 범죄 조직을 잡기 위해 ...
3	아바타: 물의 길 (부가영상 제공)	8.6	제임스 카메론	조 샐다나 샘 워싱턴 시고니 워버 우나 채플린 지오바니 리비시 스티븐 랭 케이트 ...	2022.12.14.	<아바타: 물의 길>은 판도라 행성에서 '제이크 설리'와 '네이티리'가 이룬 가족이...
4	올빼미	9.3	안태진	류준열 유해진 최무성 조성하 박명훈 김성철 안은진 조윤서 이주원 김예은	2022.11.23.	맹인이지만 뛰어난 침술 실력을 지닌 '경수'는 어의 '이형익'에게 그 재주를 인정받아...



내용

범죄도시3 대체불가 괴물형사 마석도 서울 광수대로 발탁! 베트남 납치 살해범 검거 ...

엘리멘탈(패키지: 자막판+더빙판+부가영상+코멘터리) 불 물 공기 흙 4개의 원소들이...

공조2: 인터내셔널 공조 이즈 백! 이번엔 삼각 공조다! 남한으로 숨어든 글로벌 범죄...

아바타: 물의 길(부가영상 제공) <아바타: 물의 길>은 판도라 행성에서 '제이크 ...

올빼미 맹인이지만 뛰어난 침술 실력을 지닌 '경수'는 어의 '이형익'에게 그 재주를 ...

영화의 제목과 줄거리를 합쳐서 내용 열 추가

how

감정분석

```
def senti1(i):  
    okt=Okt()  
  
    idx=DF[DF['제목']==i].index[0]  
    a=DF[DF['내용']==i][idx]  
  
    try:  
        t=okt.normalize(a)  
        score=0  
        for w,p in zip(senti['word'], senti['polarity']):  
            score=score+t.count(w)*p  
    except:  
        score=0  
  
    try:  
        if score>0:  
            rst='긍정'  
        elif score<0:  
            rst='부정'  
        else:  
            rst='중립'  
    except:  
        rst='중립'  
  
    return score, rst
```

	word	polarity
0	——	-1
1	πππ	-1
2	π_π	-1
3	π	-1
4	┐—	-1
...
14849	(^_^)	1
14850	(;_)	-1
14851	(-_-)	-1
14852	(;-)	1
14853	-_-^	-1
14854 rows × 2 columns		

Knu – senti 감성사전 사용

감성점수 부여

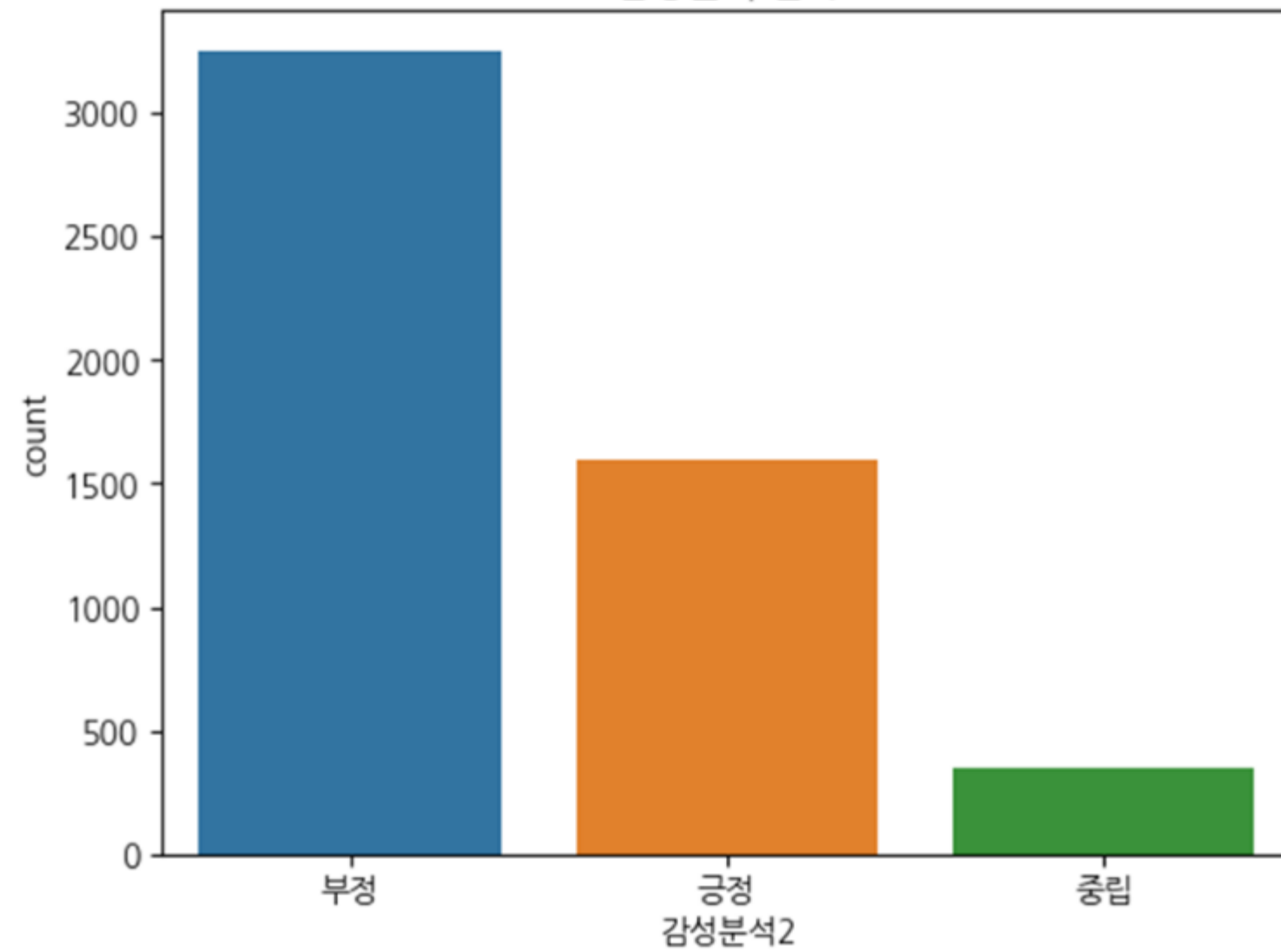
점수를 바탕으로
긍정, 부정, 중립 확인

how

감정분석 결과

내용	감성 점수	감성 분석 결과
범죄도시3 대체불가 괴물형사 마석도 서울 광수대로 발탁! 베트남 납치 살해범 검거 ...	-14	부정
엘리멘탈(패키지: 자막판+더빙 판+부가영상+코멘터리) 불 물 공기 흠 4개의 원소들이...	9	긍정
공조2: 인터내셔널 공조 이즈 백! 이번엔 삼각 공조다!남한 으로 숨어든 글로벌 범죄...	-19	부정
아바타: 물의 길(부가영상 제 공) <아바타: 물의 길>은 판 도라 행성에서 '제이크 ...	-5	부정
올빼미 맹인이지만 뛰어난 침 술 실력을 지닌 '경수'는어의 '이형익'에게 그 재주를 ...	-12	부정

감성분석 결과





how

감정분석 - 워드클라우드 결과



하늘에서 음식이 내린다면

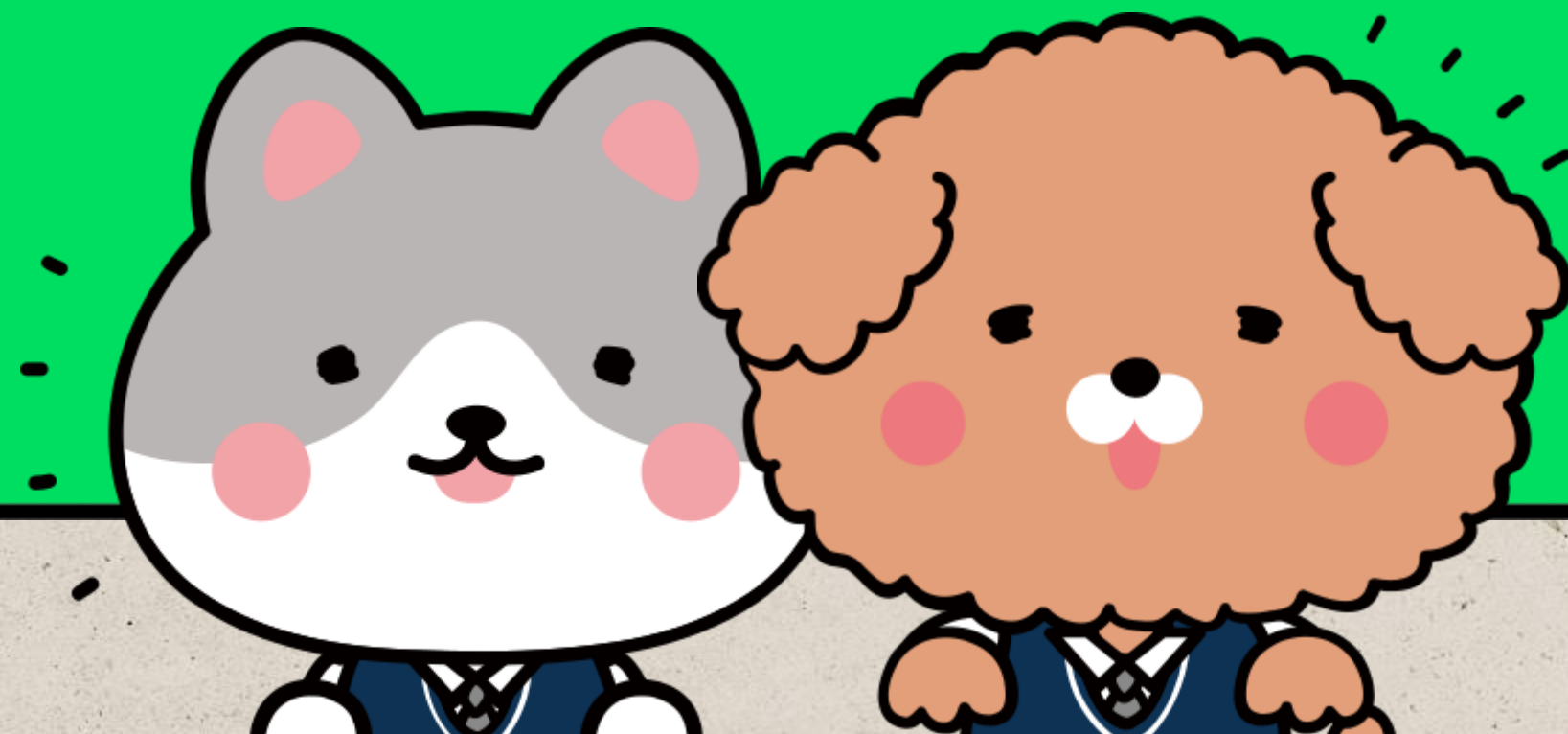


극한직업



국제시장

kkma / cosine



how

kkma

kkma를 이용하여 명사만 추출해서 csv 파일로 만들어 저장

```
import csv
# CSV 파일 경로
csv_file_path = "movie_kkma.csv"

# CSV 파일 열기 및 데이터 쓰기
with open(csv_file_path, mode='w', newline='') as file:
    writer = csv.writer(file)
    for row in total_movie:
        writer.writerow(row)

print(f"{csv_file_path} 파일에 데이터를 저장했습니다.")
```

```
from konlpy.tag import Kkma
from tqdm import tqdm

kkma = Kkma()
box = []
count = 0
total_movie = []

# 명사만 분리하여 저장함
for x in tqdm(range(0, len(df.index))):
    try:
        i = df.iloc[x]
        _name = i['제목']
        _author = i['감독'].split()
        _users = i['배우'].split()
        _description = i['줄거리']

        nouns_description_result = kkma.nouns(_description)

        total_list = []
        total_list.extend(_author)
        total_list.extend(_users)
        total_list.extend(nouns_description_result)

        total_movie.append([_name, total_list])
    except:
        pass
```

✓ 6m 42.8s

100% | 5196/5196 [06:42<00:00, 12.91it/s]

how

cosine 유사도

```
str_ = ['한국 할머니가 미국에 가서 손자 만나는']

strlist = kkma.nouns(str_[0])
str_ = [" ".join(strlist)]
print(strlist)

result = cv.transform(str_)
search_words = result.toarray()

movies_name_list = np.array(movies_name_list)

movide_df_numpy = movie_df.to_numpy()
search_words = search_words

result_cos = make_cosine(movide_df_numpy, search_words.reshape(-1))

result_args = result_cos.argsort()[::-1]
result_sorted = np.sort(result_cos)[::-1]

#결과
result_percentages = result_args
result_items = movies_name_list[result_args]

(result_sorted[:10], result_percentages, result_items[:10])
```

```
result = cv.transform(str_)

search_words = result.toarray()
```

```
total_movie
i=0
movies_name_list=[]
words_list=[]

for x in total_movie:
    i+=1
    title = x[0]
    message = x[1]
    words_list.append(" ".join(message))
    movies_name_list.append(title)

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()
words_list

mdf = cv.fit_transform(words_list)
movie_df = pd.DataFrame(mdf.toarray(), columns=cv.get_feature_names_out())
```

✓ 1.0s

how

summary

감정분석

감정사전을
불러와 적용해
줄거리 내용의
긍정과 부정을
구분해서 추가함

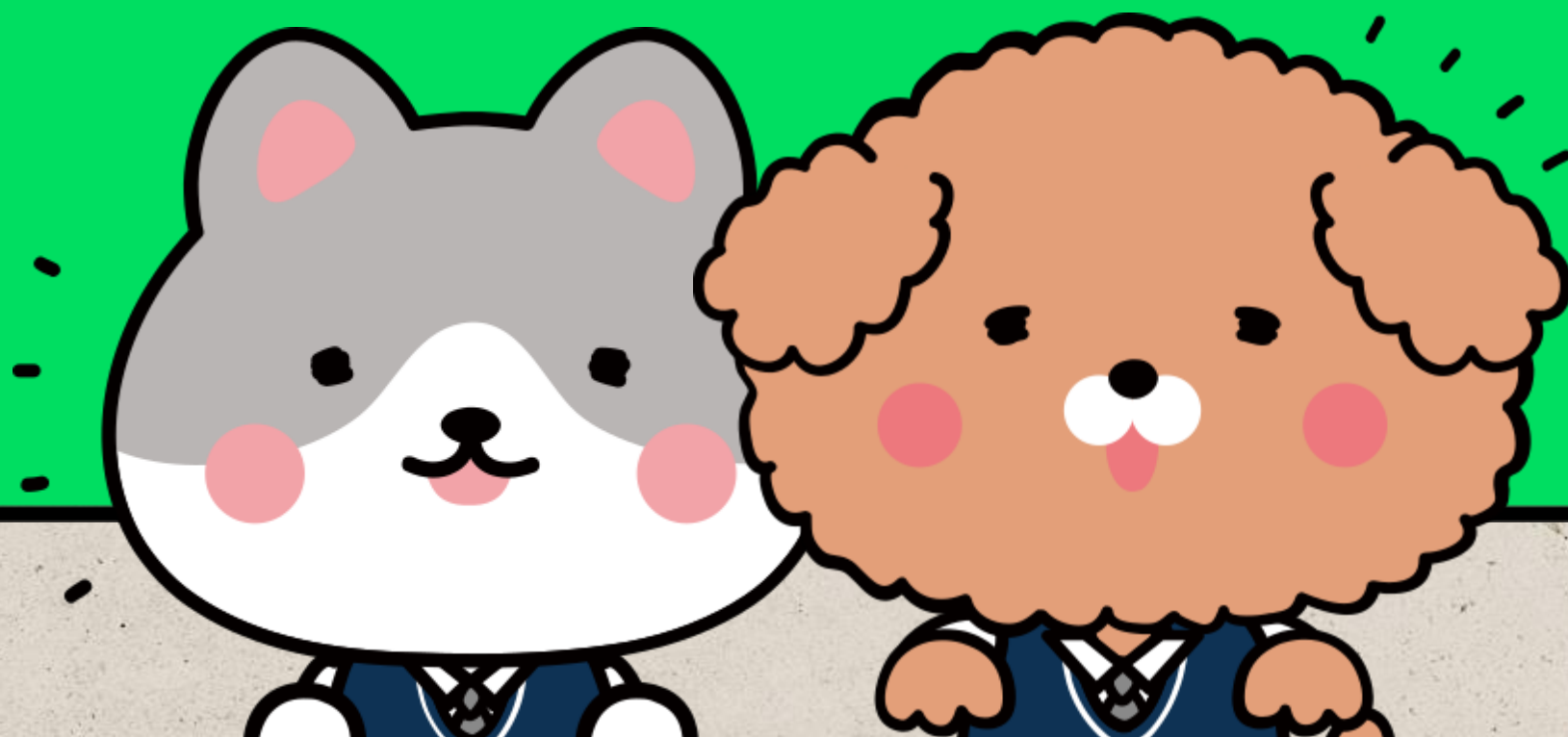
kkma

kkma를 이용해
줄거리 부분의
명사만 추출해
csv파일로 저장

cosine

cosine 처럼
-1~1사이의
유사도를 비교
값이 1에 가까
울수록 유사

유사도 분석 & 서비스 구현



Preprocess

데이터 형태 파악

제목	줄거리	명사 데이터	
범죄 **	강력 형사 **	▶	[“강력”, “형사”...]
**도시	강력 형사 **	▶	[“형사”, “강력”...]
****	감자와 고구마	▶	[“감자”, “고구마”...]

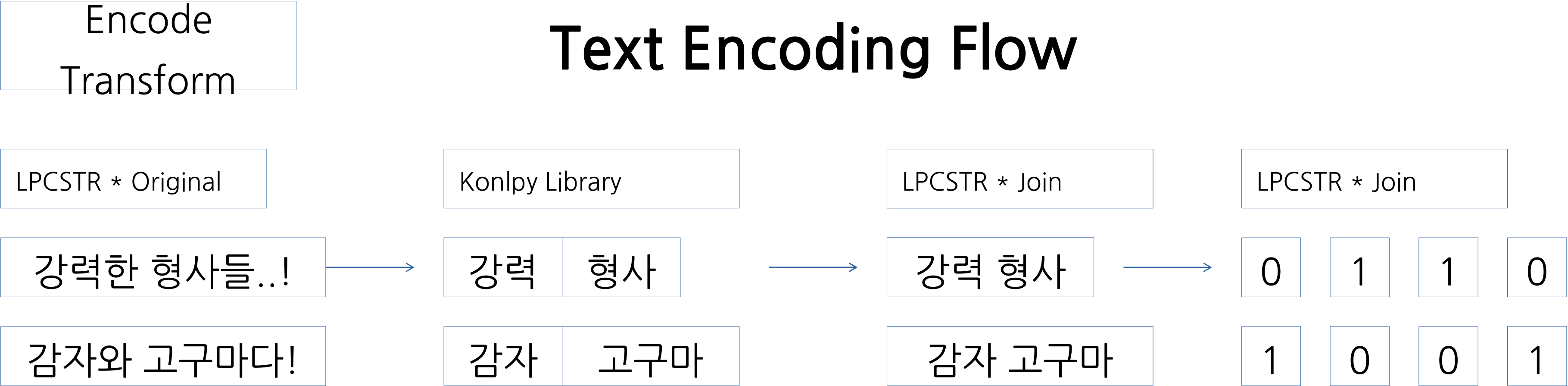
Konlpy 라이브러리

강력 형사

감자 고구마

Description Text를 Nouns 만 분할하여 새 열 정보로 기입

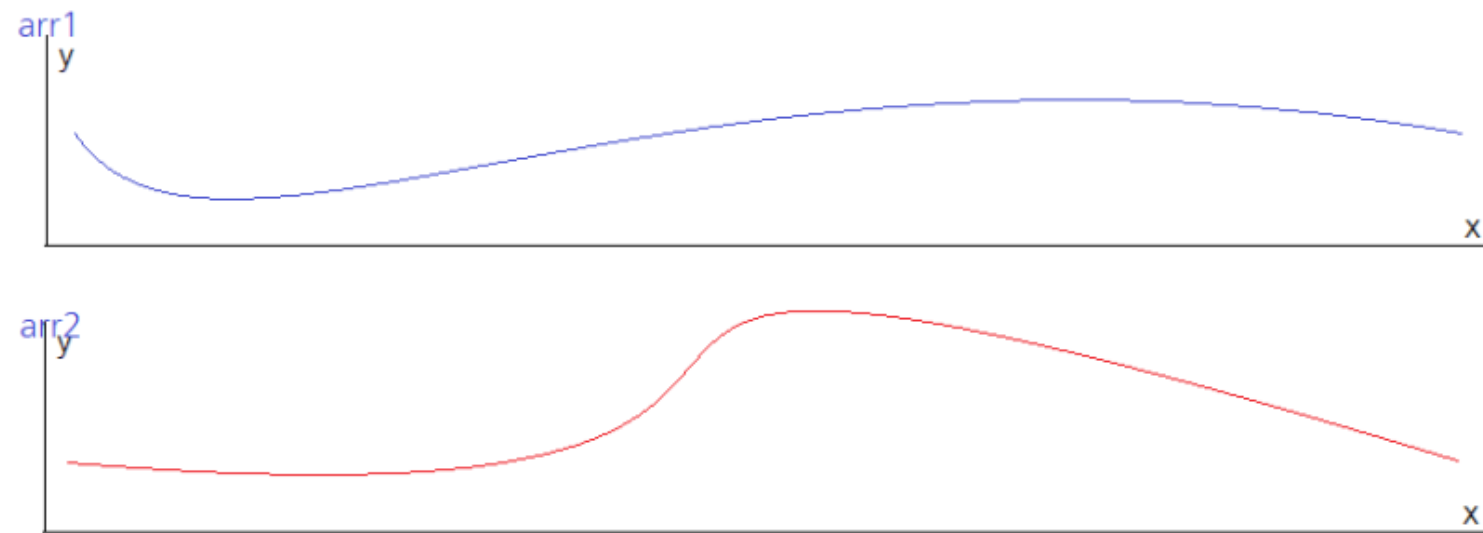
Text Encoding Flow



원본 오리지널 텍스트를 부호화(Encode)하는 과정

Data Cosine Similarity

연속 데이터의 코사인 유사도 측정



→ 그래프의 흐름 및 추이의 유사성
분석 가능

문자열 데이터의 배열 단순 비교 기법

arr1	가	나	다	라	마	바	사	아	자	차	카	파	타	하	가	나	다	라
arr2	가	나	다	라	마	바	바	아	자	차	카	바	타	하	가	나	다	라

Client Side

Index.HTML CSS

```
.top_navigator.hide{
  opacity:0;
  pointer-events:none;
}

.bottom_div{
  display:inline-block;
  position: fixed;
  bottom:0px;
  left:0px;
  width:100%;
  height:100%;

  box-shadow: 0px -2px 2px rgba(0,0,0,1);
  background: rgba(0,0,0,1);
  transition: 1s;

  z-index:50;
}

.bottom_div.bottom_div_height{
  background: rgba(30,30,30,1);

  height:200px;
}

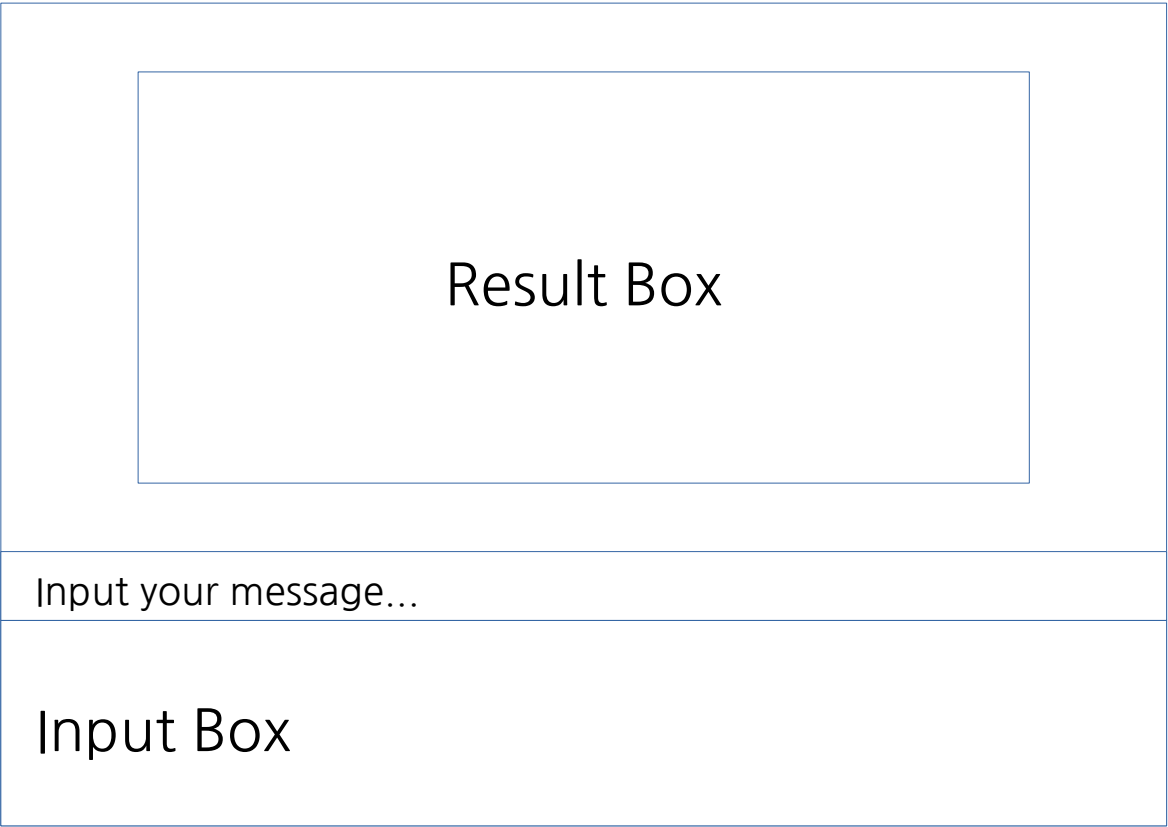
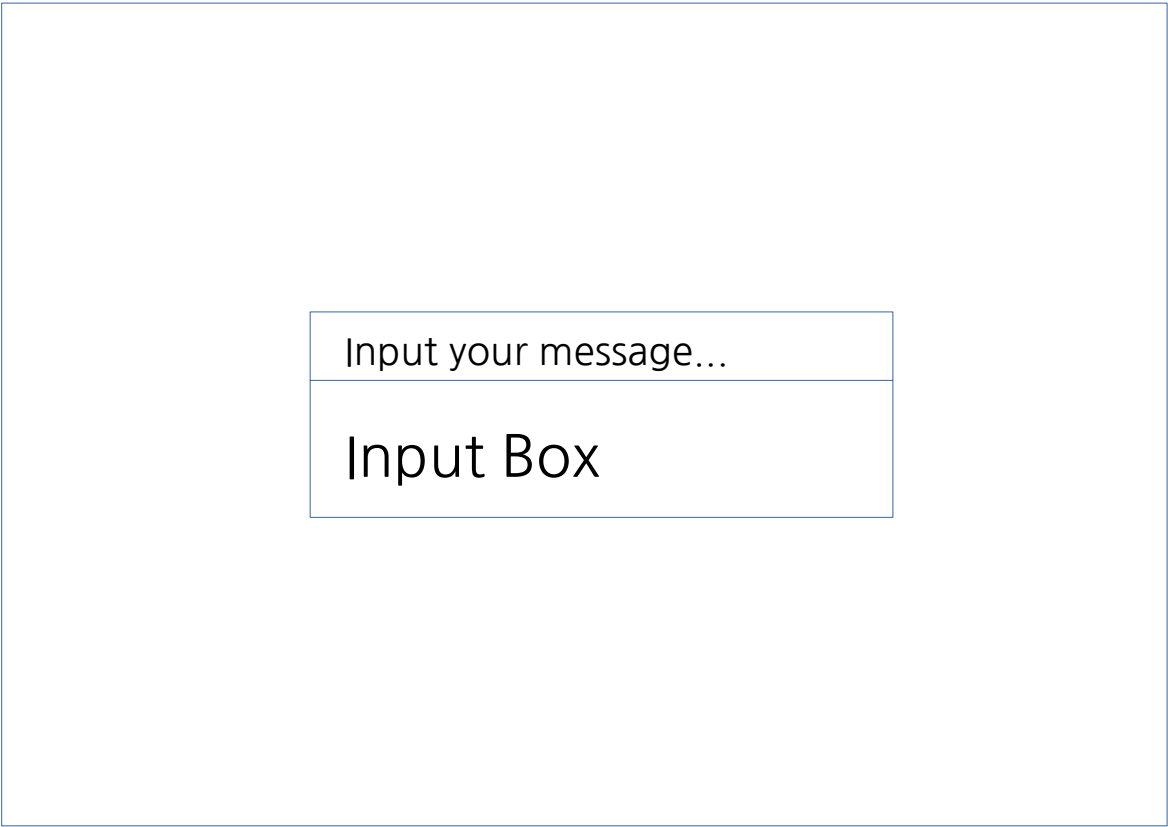
.inner_textbox{
  display:inline-block;
  position: absolute;
  top:50%;
  left:50%;

  transform: translate(-50%,-50%);

  width:70%;

  color:white;
  font-size:30px;

  transition: 1s;
}
```



↓
Flow



Animated ProgressBar

Client Side

JavaScript JS

//글자 지움 모드인 경우

```
if(now_text_index==0){  
    over_count+=1;  
    if(over_count>=10){  
  
        now_selected_text = parseInt(Math.random() * movie1.le  
  
        now_text_index=0;  
        over_count=0;  
        now_upmode=false;  
    }  
  
}else{  
    now_text_index-=1;  
}  
}
```

```
text_input_item.setAttribute("placeholder",selected_text + "|");
```

Input your message...

형..

Input your message...

형사

Input your message...

형사가 ..

...

동적 Text Animation 구성

Original Text = [“형사가 범인을 잡으려..!”]

Selected Text = “형”

Selected Text(n+1) = “형사”

Selected Text(n+2) = “형사가”

Selected Text(n+3) = “형사가 ”...

Server Side

Python Flask Library

```
@app.route("/")
def main():
    return open("works.html",encoding="UTF-8",mode="r")

@app.route("/request",methods=["GET"])
def request_function():
    content = request.args.get("content")

    _title="title"
    _author = "author"
    _actor = "actor"
    _description="description"
    _link = "link"
    _thumb = "/requestimg"

    return _title + "|" + _author+"|" + _author + "|" + _description + "|" + _link + "|" + _thumb
```

내부 서버망 게이트웨이

Server IP : 192.168.33.22

Server Gateway : 192.168.33.1

Subnet mask : 255.255.255.0

Port : 8080



But 외부망, 망에서의 접근 사용X

Connection IP/remote port : 127.0.0.1:8080

Server Side

Python Flask Library

Image Response Header

Flask 라이브러리에서 이미지 등 데이터를 전송하는 메서드/함수의 제공

```
return send_file("img.png", mimetype="image/jpeg")
```

할게 없음

MIRI.RECREATION



THE END

제목 찾아줘요! 영화 제목

조원 박소원, 배진호, 오승준, 윤성원

맡은 과정 박소원 : 크롤링, 감정분석, 워드클라우드 생성

배진호 : 코사인 유사도 분석, 서비스 구현

오승준 : 코사인 유사도 분석, PPT 제작

윤성원 : 크롤링, 감정분석