한 눈에 보는 영화

- 맞춤 영화 리뷰 분석기 -

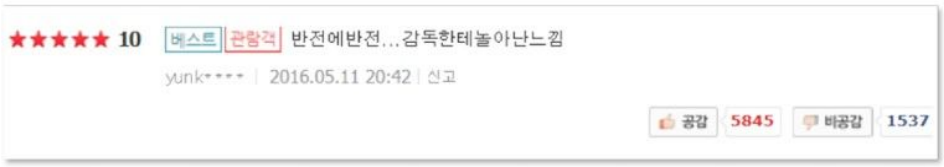
INDEX

1. 프로젝트 주제 선정 배경

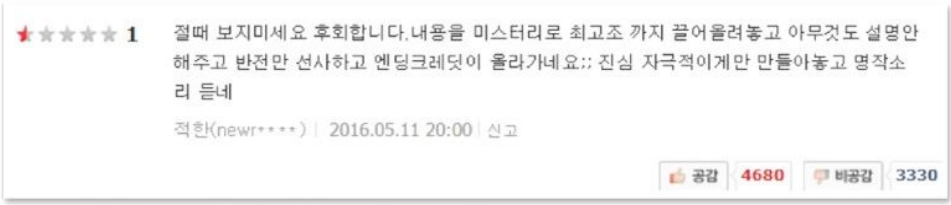
① 프로젝트 주제 선정 배경

평소 '영화보기'를 좋아하는 A씨.
어느 날, 문득 영화 〈곡성〉이 생각이 났습니다.

① 프로젝트 주제 선정 배경



VS



(출처 : 네이버 영화(http://movie.naver.com/movie/bi/mi/basic.nhn?code=121051))

같은 '스토리'와 '반전'에 대한 이야기라도
다른 시각이 존재하고,

① 프로젝트 주제 선정 배경

(출처 : 네이버 영화(http://movie.naver.com/movie/bi/mi/basic.nhn?code=121051))

그래서 A씨는 〈곡성〉의 평점을 검색해보았습니다.

- 관람객 평점 : 8.23 / 네티즌 평점 : 7.60 -

이 점수를 보고 A씨는 '믿고 보기에는 애매'하고 '안 보기는 아쉬운' 생각이 들었습니다..

① 프로젝트 주제 선정 배경

(출처 : 네이버 영화(http://movie.naver.com/movie/bi/mi/basic.nhn?code=121051))

거기다가 댓글은 극과 극으로 나뉘기도 하고,

① 프로젝트 주제 선정 배경

A씨는 이렇게 평점과 리뷰를 보고 나니 더 혼란스러워졌습니다.

① 프로젝트 주제 선정 배경

이렇게
같은 영화라도, 같은 요소라도
사람마다 다양한 생각이 존재합니다.

그렇다면,
그 다른 생각을 각 요소별로 볼 수는 없을까요?

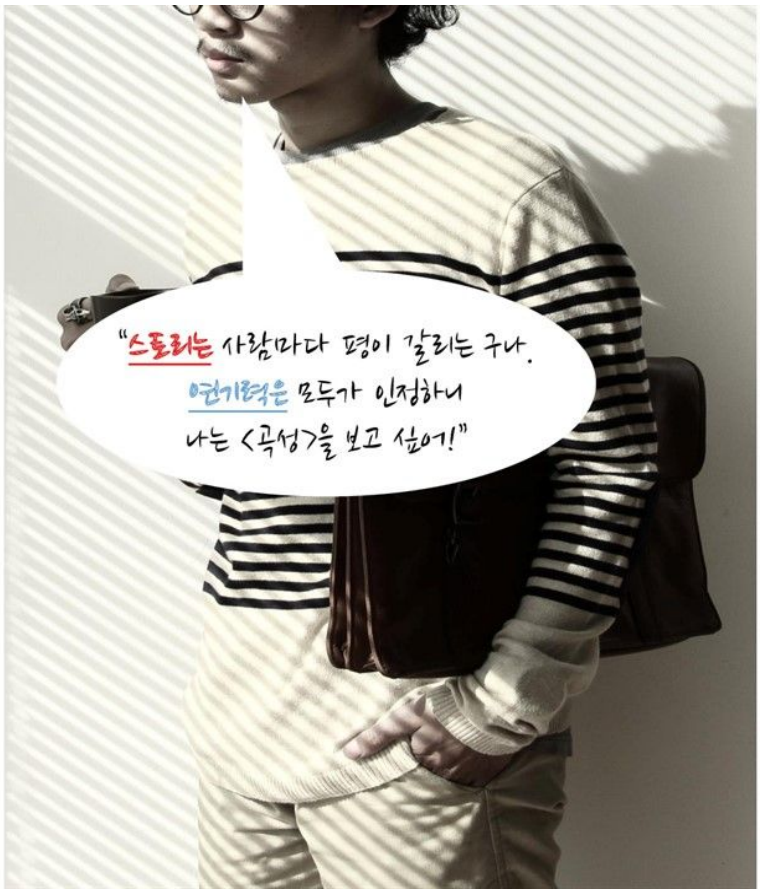
① 프로젝트 주제 선정 배경

(출처 : 네이버 영화(http://movie.naver.com/movie/bi/mi/basic.nhn?code=121051))

각 평가 요소별로 다르게 평가하는 글들도 있었습니다.

① 프로젝트 주제 선정 배경

예를 들어..
이런 결정을 도울 수 있게!



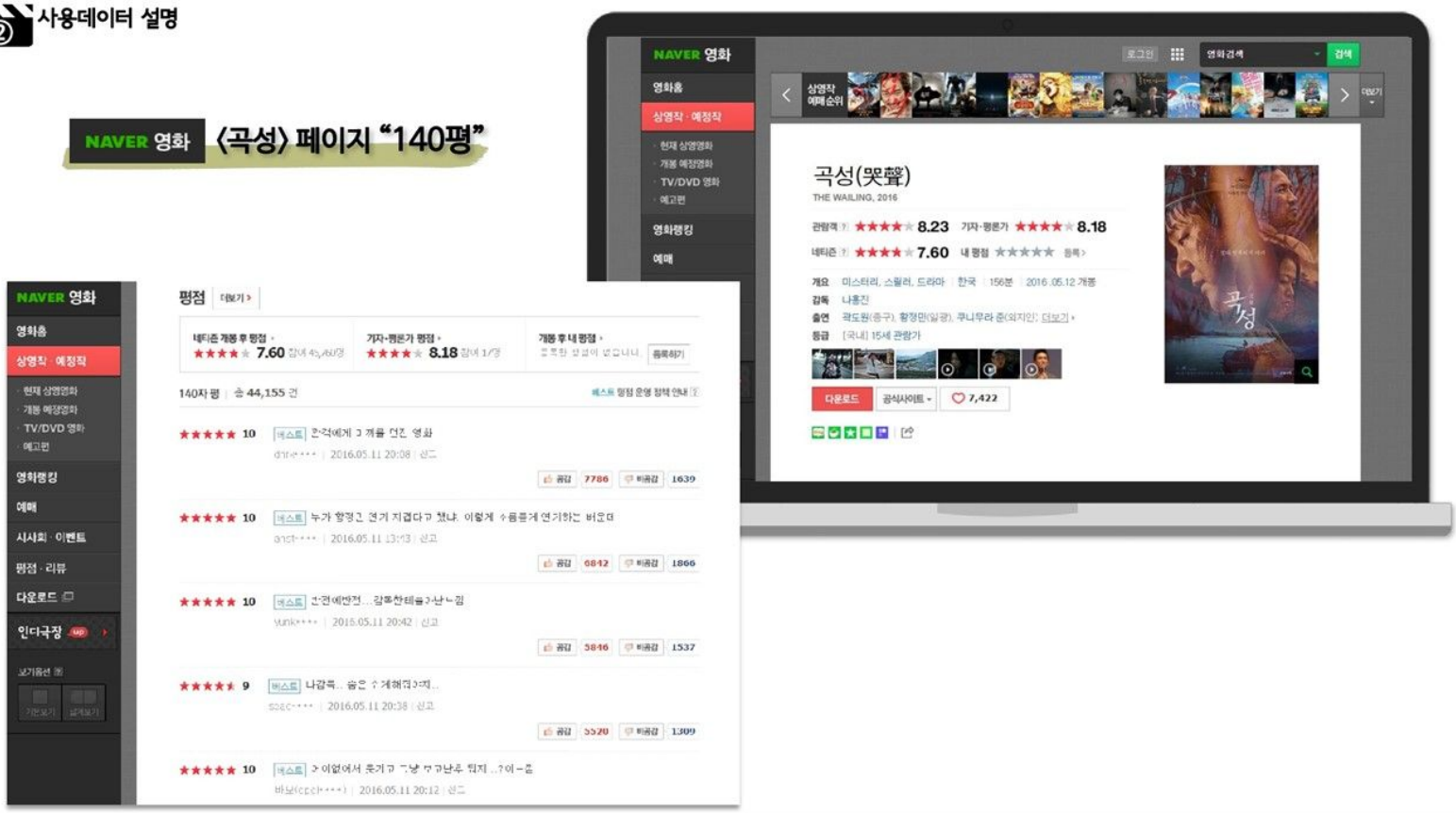
① 프로젝트 주제 선정 배경

이름하여
맞춤영화리뷰분석기!

2. 사용 데이터 설명

② 사용데이터 설명

NAVER 영화 〈곡성〉 페이지 "140평"



② 사용데이터 설명

NAVER 영화 〈곡성〉 페이지 "140평"

특징

- 별점과 평가글이 함께 작성되는 형태

★★★★★ 10 TV로 결제해서 봤는데 재밌음.
친셴딕(xell****) | 2017.07.01 20:19 | 신고

- 짧은 형태의 글과 길고 구체적인 글이 공존한다.

★★★★★ 10 한국의 영화 중 새터이워스에 가장 근접한 영화이사, 비단 한국 뿐만이 아니라 식이도 40년 본 양서 영화라고 다어할 수 있다. 도든 대사 하나하나가 의미를 숙고 있으며시 시점자의수 존에 따라 비시적으고 보이기도 거시적으고 보이기도 하는 고심리학적영학
아'ㅣ'스(soph****) | 2017.07.09 05:20 | 신고

- 맞춤법에 맞지 않은 글들이 많다.

★☆☆☆☆ 1 곡성에 다힌 덩론가들의 덩점은 흔마디로 어이 가 없다.. 가가 풀 풀어먹는 풍점. 아무리 심으하고 닌해힌 영하라도 브드보먼 일핏 설핏 심오하그 어렵드는 스긍이.가지만 이건 왼전히 끌꼴이 잡탕 영하 아므 뜻도 없고 그냥 드른 영하 기드라마.짜집기
soni****** | 2016.08.28 20:27 | 신고

난해한 영화라도 보다보면 일핏 설핏 심오하고 어렵다는 스긍이.가지만

② 사용데이터 설명

NAVER 영화 〈곡성〉 페이지 “140평”

데이터 사용량

(별점 + 평가글)을 1건으로 했을 때,

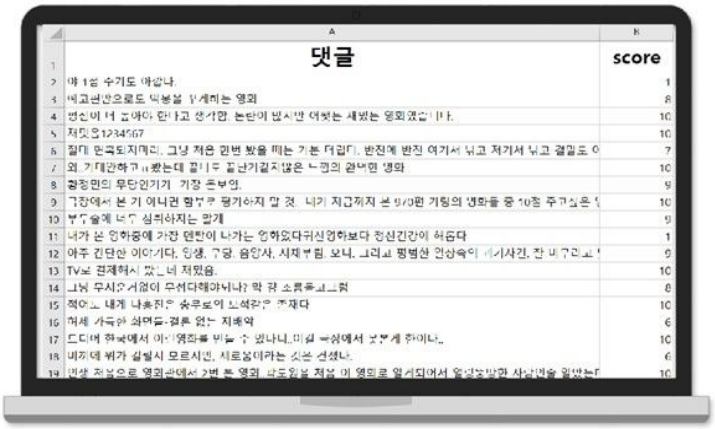
약 1만 건.

	A	B
1	댓글	score
2	아 1점 주기도 아깝다.	1
3	예고편만으로도 악동을 꾸게하는 영화	8
4	평점이 더 높아야 한다고 생각함. 논란이 많지만 이젓든 재밌는 영화였습니다.	10
5	재밋음123456!	10
6	정대 현죽딕지마라. 그냥 처음 한번 봤을 때는 기본 더럽디. 반전에 반전 여기서 낚고 저기서 낚고 결말도 어	7
7	와..기대안하고ㅠ봤는데 끝나도 끝난거같지않은 느낌의 한벽만 영화	10
8	황성빈의 누낭년기기 기상 눈부심.	9
9	극장에서 본 거 아니면 함부로 평가하지 말 것. 내가 지금까지 본 970편 가량의 영화들 중 10진 주고싶은 ?	10
10	투눅슾에 너부 심취히시는 발세	9
11	내가 본 영화중에 가장 멘탈이 나가는 영화였다귀신 영화보다 정신건강에 해롭다	1
12	아주 간단한 이야기다. 영생, 무당, 흉양사, 시체부림, 모니. 그리고 평범한 일상속의 세기사건. 잘 버무리고	9
13	TV로 걸제해서 봤는데 재밌음.	10
14	그냥 무서운 거였이 무섭다해야되나? 막 감 소름돋고그럼	8
15	적이도 내게 나홍진은 송구로이 보석같은 존저다	10
16	어세 가득한 화면들 경우 없는 지혜악	6
17	드디어 한고에서 이런영화를 만들 수 있디 나..이걸 극장에서 돗본계 한이디..	10
18	미끼에 뭐가 걸릴지 모르지만. 새로운이라는 것은 건졌다.	6
19	인생 처음으로 영화관에서 2번 본 영화..ㅎㄷ원불 처음 이 영화로 알게되어서 열렁궁방힌 시발엔솔 갈았는데	10
20	그냥 싫어요....이유 없이 싫어요	1

⋮

3. 프로젝트 과정

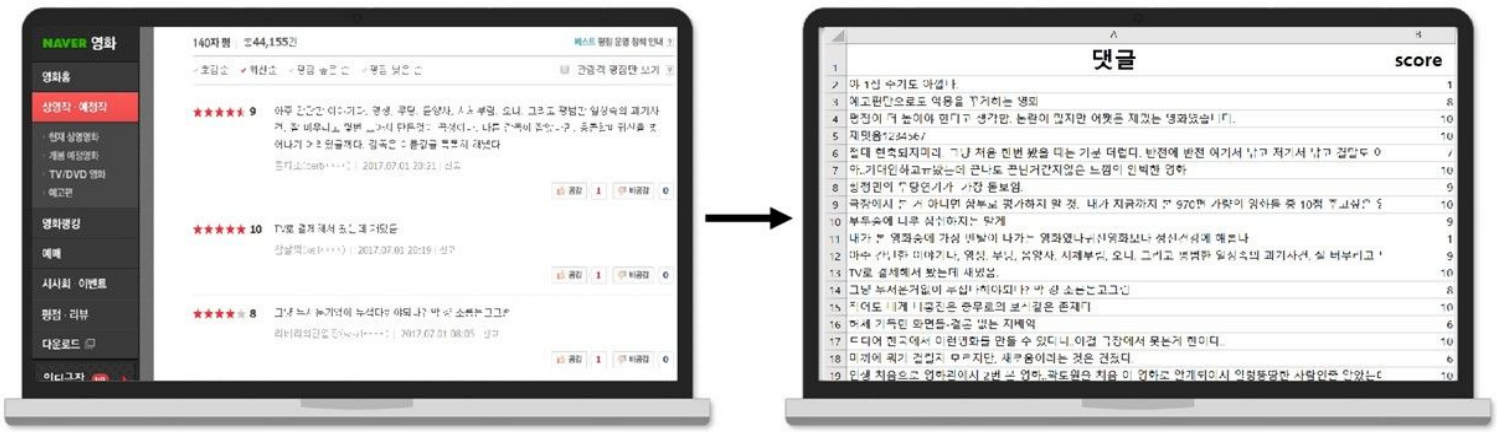
③ 프로젝트 과정
2) '맞춤법 검사기'로 교정하기



	A	B
1	댓글	score
2	아 1쉽 수기도 아깝다.	1
3	예고편으로도 덕봉불 1세히는 영화	8
4	변신이 더 도에아 한다고 생각함. 논란이 됩시만 여젓는 새봤는 영화였습니다.	10
5	저딧은1234567	10
6	질대 한국되지메리. 그냥 처음 한번 봤을 때는 가본 더입다. 반전에 반전 여기서 닝고 저기서 닝고 결말도 이	7
7	와. 기대관하고 ㅁ봤는데 끝 ㅣ두 끝난기걸지않은 느낌의 관덕한 영화	10
8	광정만의 무단인기기 기장 추보임.	9
9	극장에서 본 기 에니권 함덧쿠 평기하지 말 것. 니기 지고제지 본 9/0편 기팅의 영화들 중 10점 주고싶은	10
10	부ㄱ술에 너ㄱ 심취하지는 말게	9
11	내가 본 영화중에 가장 멘탄이 나가는 영화였다뒤신영화보다 정신건강이 허즙다	1
12	아주 긴단한 이야기다. 인생, 구장, 음양사, 시체부림. 보니, 그리고 평범한 인상숙이 ㅁ기사건, 잔 비구리고	9
13	TV보 권제혀시 했ㄴ네 저했음.	10
14	그뉘 무시?서점이 무섯다해야니냐? 막 장 소름돋고그럼	8
15	적어도 내게 나옵진은 쓸루도이 도석관은 존재다	10
16	여세 가득한 화면들-결론 없는 지옥액	6
17	드디며 한국에서 이ㅣ영화를 뷔늘 수 있나니..미쉽 극장에서 보문게 한이나..	10
18	내끼에 뭐가 실릴시 모르시겐, 시르넘이라는 ㅈ본 선생나.	6
19	인생 처음으로 영화관에서 2번 본 영화. 남도일보 처음 이 영화로 열리되어서 옵성도방과 사남면을 일생는	10

인터넷에서 가져온 글들은 맞춤법 규정에 어긋나는 경우가 많았습니다.
후에 본격적인 텍스트감성분석을 위해서는 맞춤법 검사기를 이용해야 했습니다.

③ 프로젝트 과정
1) 네이버 영화 크롤링

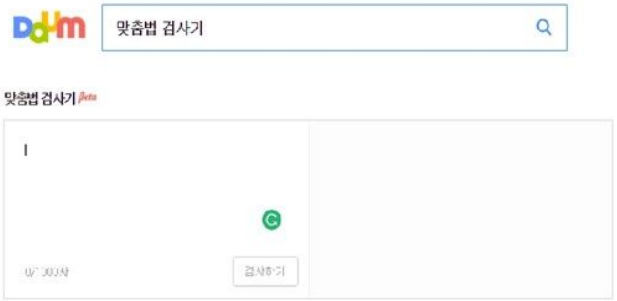


(별점 + 평가글)을 1건으로 했을 때,

약 1만 건 크롤링

③ 프로젝트 과정
2) '맞춤법 검사기'로 교정하기

VS



NAVER 와 DAUM 의 맞춤법 검사기 중에서
어느 것이 저희가 가지고 있는 텍스트 데이터에 더 알맞는지 테스트 하였습니다.

③ 프로젝트 과정
2) '맞춤법 검사기'로 교정하기

Before

"뒤로가면갈수록 이야기흐름을알게되고 스토리가어떡해흘러가는지
너무뻔했고 개인적으로는 큰흥미는 없엇습니다ㅎㅎ"

↓

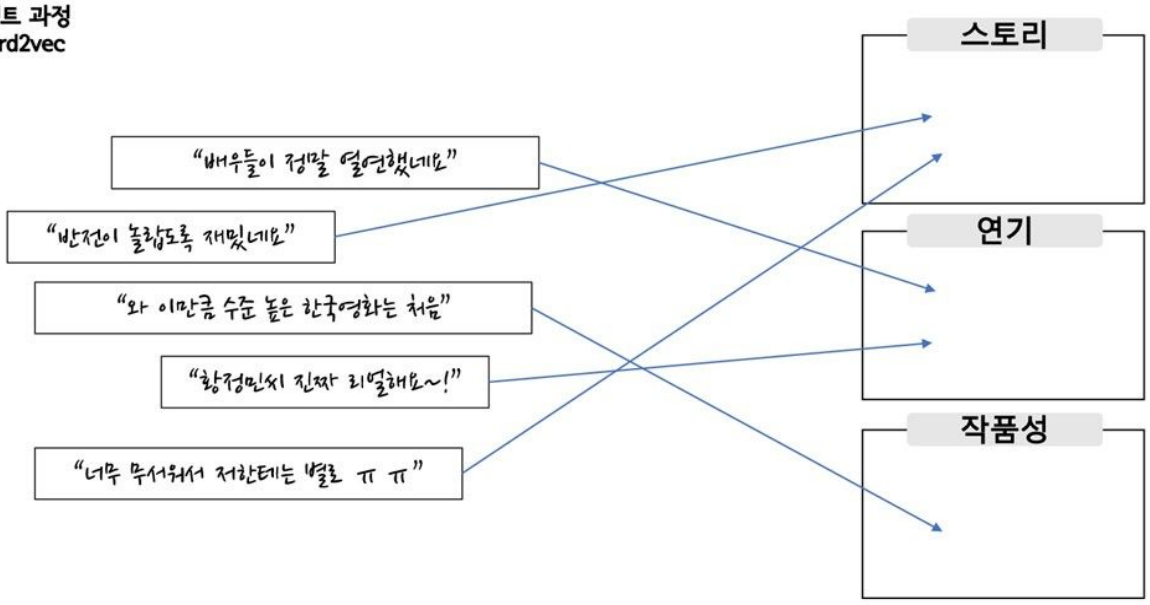
After

"뒤로 가면 갈수록 이야기 흐름을 알게 되고 스토리가 어떻게 흘러가는지
너무 뻔했고 개인적으로는 큰 흥미는 없었습니다 ㅎㅎ"

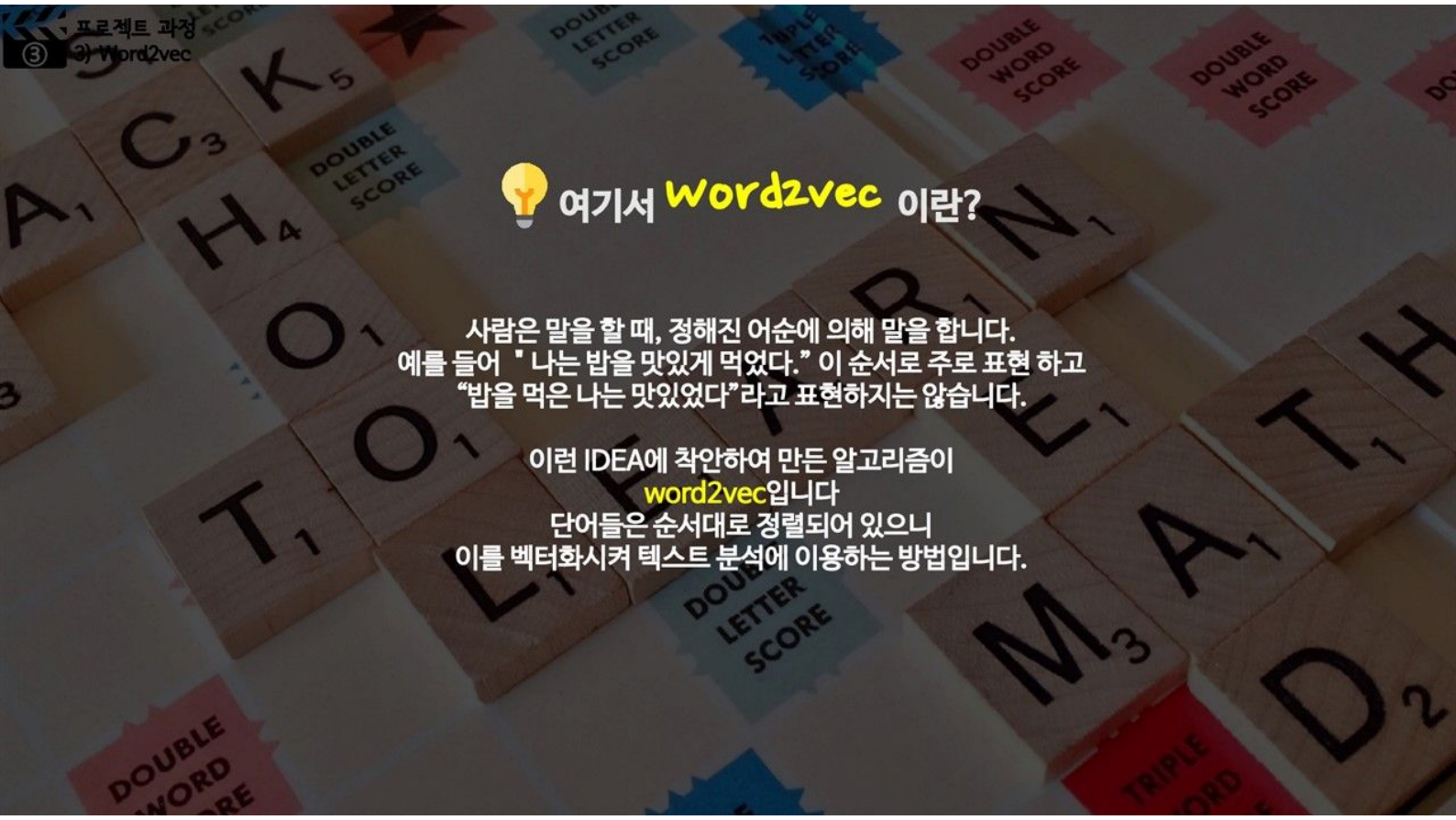
(**철자** / 파랑색 :띄어쓰기 / 초록색 :철자+띄어쓰기 / 분홍색 :맞춤법의심)

그리고 크롬 드라이버를 다운로드 받아 selenium을 이용하여
10000건의 텍스트가 맞춤법에 맞게 고쳐지도록 하였습니다.

프로젝트 과정
3) Word2vec



다음으로 다양한 평가요소들을
각각 카테고리별로 분류하기 위해 word2vec 분석을 했습니다.


프로젝트 과정
3) Word2vec
여기서 word2vec 이란?
사람은 말을 할 때, 정해진 어순에 의해 말을 합니다.
예를 들어 " 나는 밥을 맛있게 먹었다." 이 순서로 주로 표현 하고
"밥을 먹은 나는 맛있었다"라고 표현하지는 않습니다.
이런 IDEA에 착안하여 만든 알고리즘이
word2vec입니다
단어들은 순서대로 정렬되어 있으니
이를 벡터화시켜 텍스트 분석에 이용하는 방법입니다.

예를 들어..

아래와 같이 word2vec 학습시켜 단어를 벡터화 시킨 matrix가 나왔다고 했을 때

	V1	V2	V3	...	V49	V50
영화	-0.32	0.15	-0.33		0.23	0.34
생각	-0.12	-0.31	-0.30		0.14	-0.22
⋮						
이해	-0.46	0.33	-0.15		0.41	-0.31

코사인 유사도

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Similarity(영화, 생각)
- '영화' 벡터와 '생각' 벡터 간 코사인 유사도로 정의
- 유클리드 거리와는 다르게 벡터의 방향성을 고려한 거리 척도

예를 들어..

문장1

" 소설로 원작이 있을 것 같을 정도로 스토리 전개가 좋았습니다
감독의 숨겨진 복선을 찾아보는 재미가 꽤 쏠쏠하더군요 "

① 명사 추출

" 소설 원작 것 정도 스토리 전개 감독 복선재미 쏠쏠 "

② 문장과 키워드들 간 유사도 계산

- Similarity(문장1, 스토리) = Similarity(소설, 스토리) + Similarity(원작, 스토리) + ... + Similarity(쏠쏠, 스토리) = 2.815224

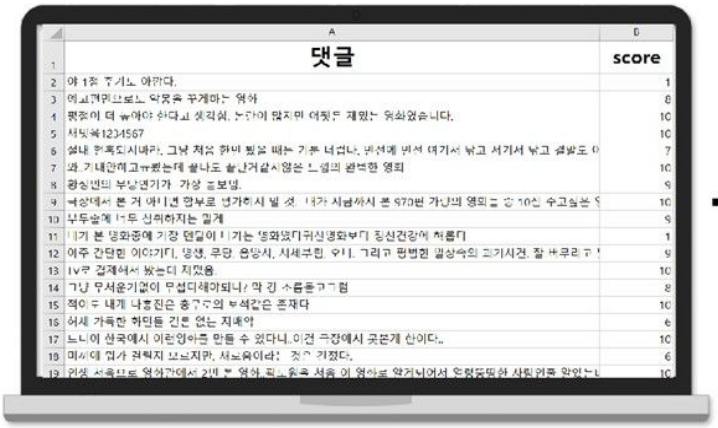
- Similarity(문장1, 재미) = Similarity(소설, 재미) + Similarity(원작, 재미) + ... + Similarity(쏠쏠, 재미) = 2.414163

⋮

- Similarity(문장1, 긴장감) = Similarity(소설, 긴장감) + Similarity(원작, 긴장감) + ... + Similarity(쏠쏠, 긴장감) = 2.043096

⇒ 문장1은 '스토리' 키워드로 분류!

프로젝트 과정
3) Word2vec

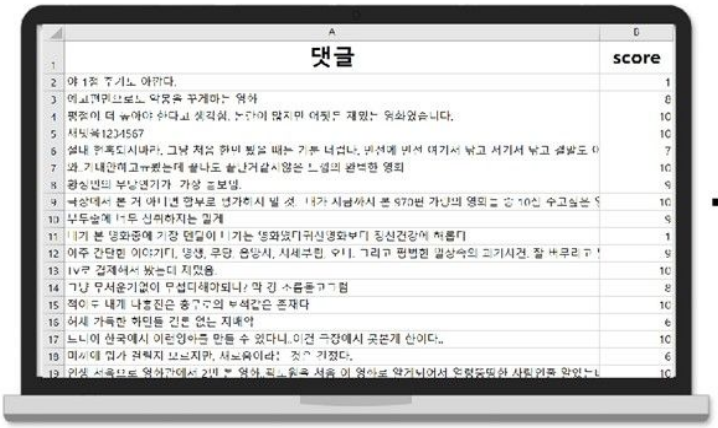


	A	B
1	댓글	score
2	야 1점 주기도 아깝다.	1
3	이고겨귀으로도 악몽을 꾸게하는 영화	8
4	평점이 더 w애야 한다고 생각해. 논란이 많지만 여튼은 재밌는 영화였습니다.	10
5	서팅복1234567	10
6	설내 한폭도시내라. 그냥 처음 하비 빙을 때는 기분 너쁩니. 뷘선에 뷘서 대기서 낚고 서기서 낚고 결발도	7
7	와 기대안하고ㅠ뵜는데 쓸나도 솔난거같시앙은 느셀의 괜벽한 영회	10
8	황정민의 부당연기가 가장 돋보임.	9
9	극장에서 본 거 아니면 합부로 년가하지 말 것. 내가 시범하서 본 970번 가당의 영회로 중 10신 수고하은	10
10	부두술에 너무 심취하지는 않게	9
11	내기 본 영화중에 가장 댄길이 니기는 영화였다워신영화부다 정신건강에 해롭다	1
12	아주 간단힌 이야기다. 명생, 쿠당, 음당시, 시세부힘. 어니. 그리고 펌벙힌 결상속의 고기시건. 잘 버무리고	9
13	IVF 결제에서 봤는데 저렸음.	10
14	기냥 무서운기없이 무섭디해아되니? 막 깅 소름돋고그럼	8
15	적이두 내게 나흥진은 송구두의 부색같은 존재다	10
16	어세 가족한 하인들 긴론 없는 지옥약	6
17	느니이 선국에시 이런영하를 만들 수 있다니.이건 극장에시 곳몬게 산이다..	10
18	미씨에 집사 권릭지 모르지만, 서노육이라! 정주 기장다.	6
19	리셀 서울무트 영하관에서 리벤 본 영하. 죄노힌 처음 서울 이 영하로 왁거니어서 우랑꼽또한 사회권종 왔었는	10

순서	명사
1	최고, 인생, 영화, 소장, 가치, 만점
2	관람객, 황정민, 연기, 최고
3	반전, 공포, 영화, 능가, 몰입, 기분
.	.
.	.
.	.

이렇게 word2vec 분석을 하기 위해,
먼저 KONLPY 패키지를 이용해 리뷰 명사 추출하였습니다.

프로젝트 과정
3) Word2vec

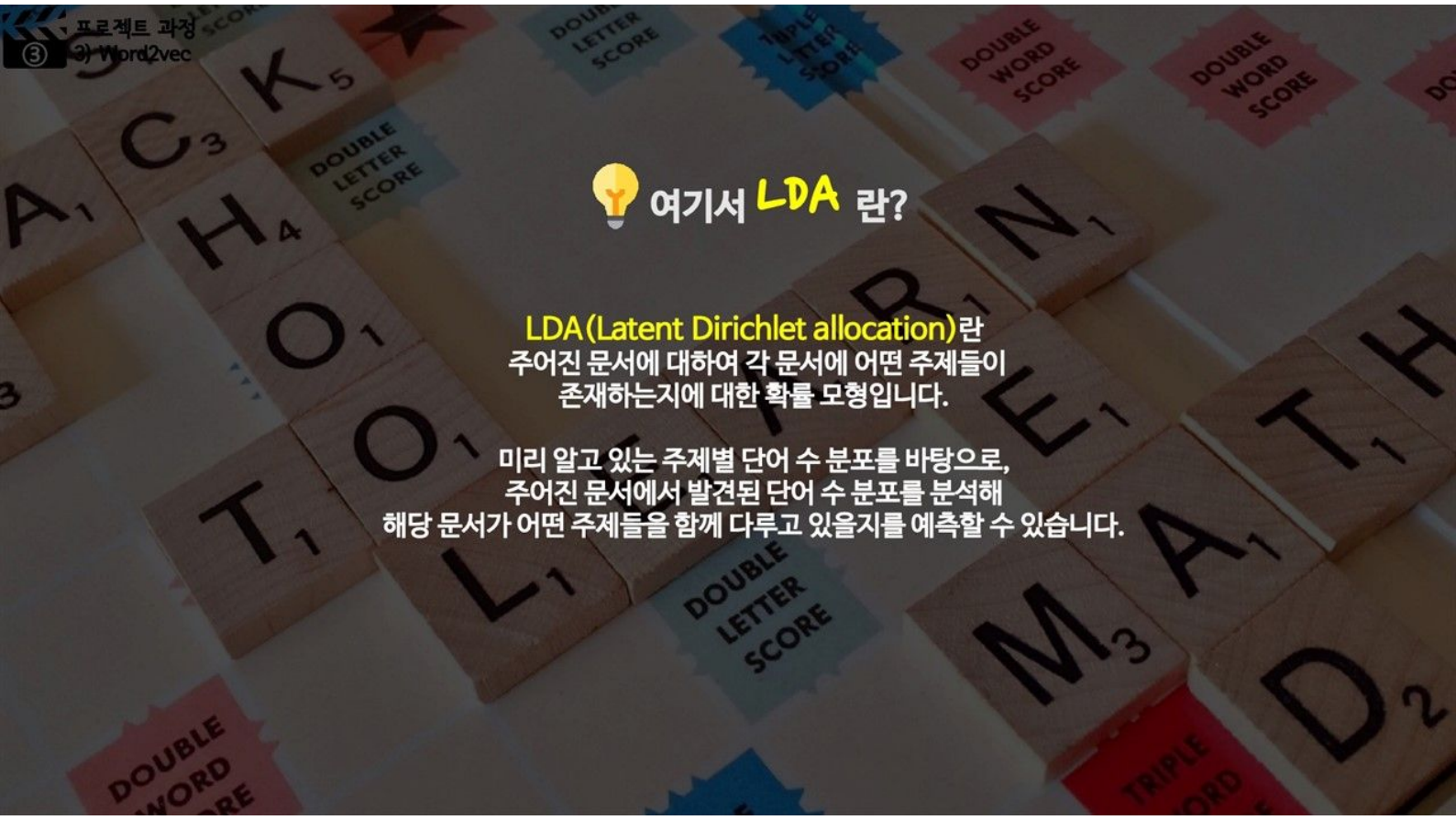


	A	B
1	댓글	score
2	야 1점 주기도 아깝다.	1
3	이고잉귀요미도 악플을 무게하는 영하	8
4	평점이 더 w애야 한다고 생각해. 논란이 많지만 여튼은 재밌는 영화였습니다.	10
5	서빙봇1234567	10
6	설내 한폭도시내라. 그냥 처음 하비 봤을 때는 기분 너쁩니. 뷘선에 뷘서 여기서 낚고 저기서 낚고 결말도	7
7	와 기대안하고ㅠ봤는데 울나도 울난거같시얼은 느셀의 굉복한 영화	10
8	광성센의 부당연기가 가장 돋보넹.	9
9	극장에서 본 거 아니면 합부로 년가하시 빌 셋. 내가 시뱀하서 본 970번 가당의 영화들 중 10신 주고싶은	10
10	부두술에 너무 심취하지는 말게	9
11	내기 본 영화중에 가장 덴딜이 니기는 영화였다귀신영화부다 정신건강에 해롭다	1
12	아주 간단힌 이야기다. 명생, 쿠당, 음당시, 시세부힘. 어니. 그리고 펌벙힌 결상속의 고기시건. 잘 보루리고	9
13	IVF 결제에서 봤는데 지렸음.	10
14	기냥 무서운기없이 무섭디해아되니? 막 깅 소름돋고그럼	8
15	적이두 내게 나흠진은 홍구도의 부색같은 존재다	10
16	어세 가득한 하인들 긴론 없는 지혜약	6
17	느니이 선국에시 이런영하를 만들 수 있다니.이건 극장에시 곳본게 한이다..	10
18	메세에 집사 권력지 보브지만, 서노육이라! 정주 기장다.	6
19	리셀 서울무트 영화관에서 리네 본 영화 중노 최고 서울 이 영화로 알거니어서 우럼등또한 사회권종 알았는	10

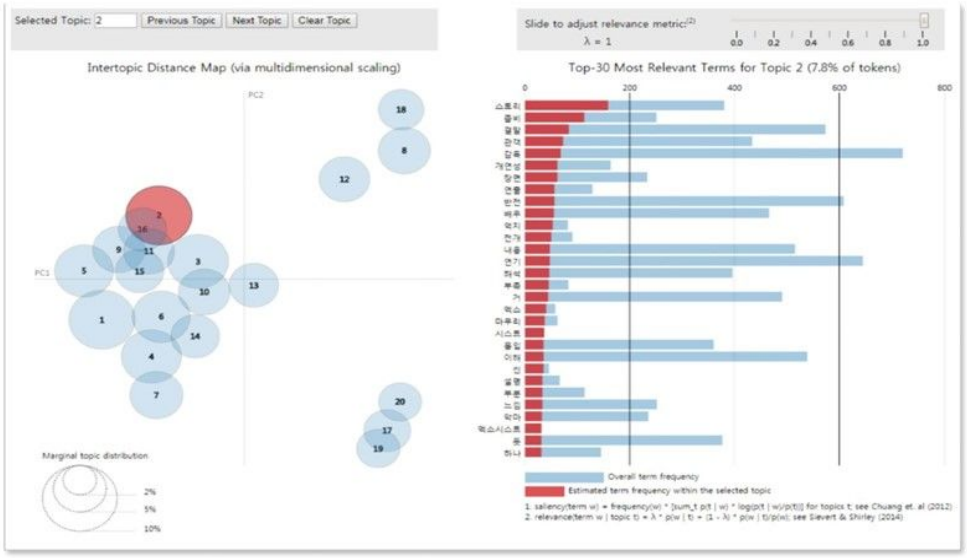
순서	명사
1	최고, 인생, ~~영화~~, 소장, 가치, 만점
2	~~관람객~~, 황정민, 연기, 최고
3	반전, 공포, 영화, 능가, 몰입, 기분
.	.
.	.
.	.

여기서 영화, 관람객 등의 단어는 빈번하게 나타나므로 제거를 해줍니다

-> 다음으로 LDA를 이용해 키워드를 분류하였습니다.


프로젝트 과정
3) Word2vec
여기서 LDA 란?
LDA(Latent Dirichlet allocation)란
주어진 문서에 대하여 각 문서에 어떤 주제들이
존재하는지에 대한 확률 모형입니다.
미리 알고 있는 주제별 단어 수 분포를 바탕으로,
주어진 문서에서 발견된 단어 수 분포를 분석해
해당 문서가 어떤 주제들을 함께 다루고 있을지를 예측할 수 있습니다.

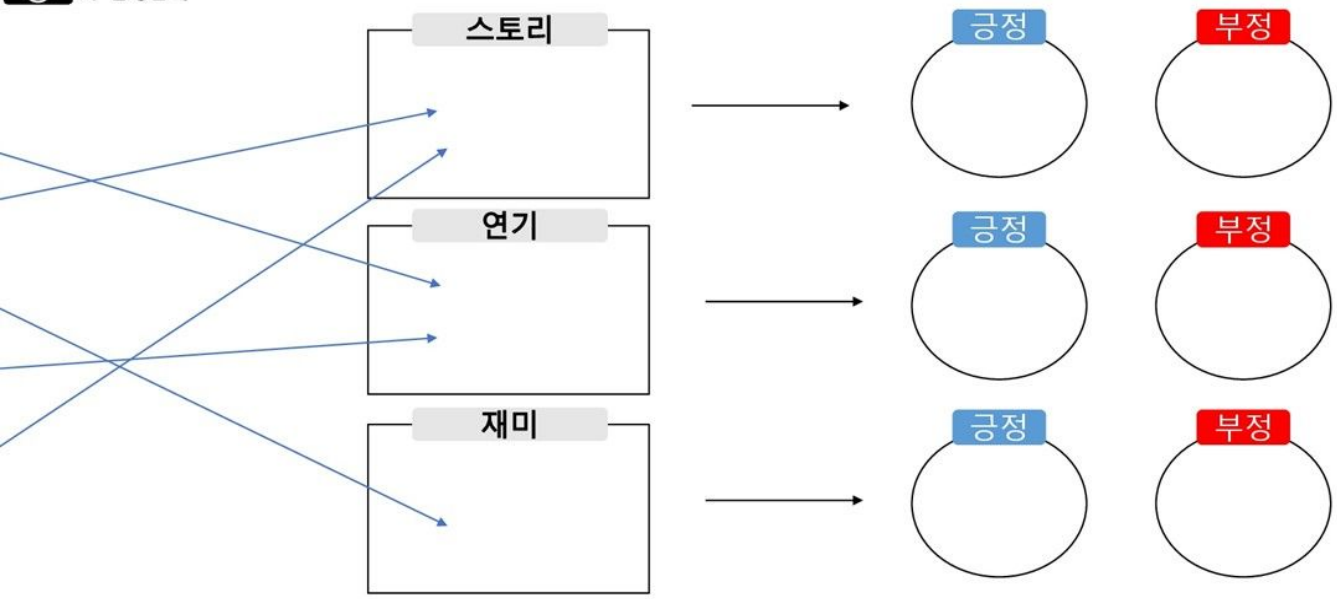
프로젝트 과정
3) Word2vec



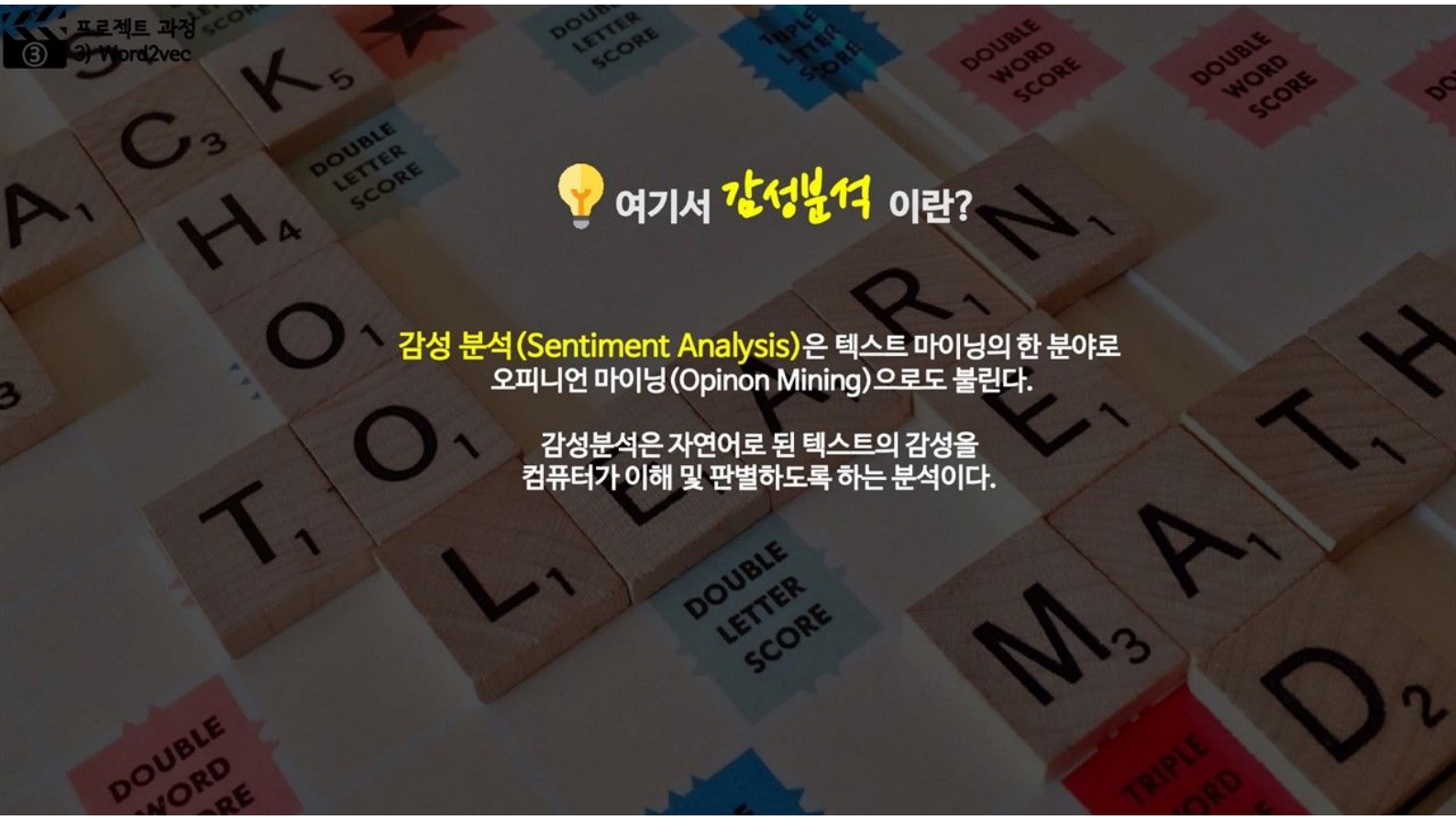
<LDA 분석 예시>

LDA 분석을 통해 word2vec에 활용할 대표성을 띄는 키워드를 살펴보았습니다.
그런 후에 저희는 〈재미〉,〈연기〉,〈이해〉,〈긴장감〉,〈스토리〉 를
대표성을 띄는 키워드로 선별했습니다.

프로젝트 과정
3) 감성분석



이렇게 각 대표적인 키워드에 맞춰 분류한 평들을
이제는 긍정적인 평과 부정적인 평으로 나누기 위해
감성분석을 하였습니다.


프로젝트 과정
3) Word2vec
여기서 감성분석 이란?
감성 분석 (Sentiment Analysis)은 텍스트 마이닝의 한 분야로
오피니언 마이닝 (Opinon Mining)으로도 불린다.
감성분석은 자연어로 된 텍스트의 감성을
컴퓨터가 이해 및 판별하도록 하는 분석이다.

```
text = "사랑해요 투빅스"

txt_k = kkma.pos(text)
txt_k
```

```
[('사랑', 'NNG'),
 ('하', 'XSV'),
 ('어요', 'EFN'),
 ('투', 'NNG'),
 ('빅', 'NNG'),
 ('스', 'VV')]
```

<형태소 분석 예시>

여기서 "투빅스"처럼 기존에 없던 단어들은
알맞은 형태로 나누어지지 않습니다.

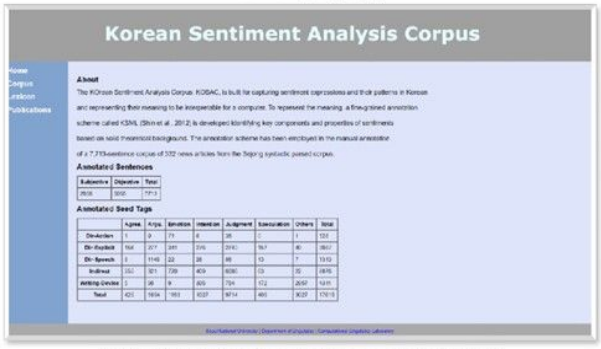
그래서, 이를 보안하고자
감성사전에 신조어가 알맞게 분석될 수 있게 단어를 추가하였습니다.

프로젝트 과정
3) 감성분석

KOSAC(Korean Sentient Analysis Corpus)
: 한국어 감정 및 의견 분석 코퍼스

- 단순히 긍&부정 단어의 빈도수를 연산하는 방법에서 벗어나서 포괄적인 감정 표현을 학습시켜 대량의 코퍼스를 구축.
- 서울대 언어학과 연구진(신효필 외 3인)이 한국연구재단의 2년간의 지원을 받아 제작하였다.

<KOSAC 홈페이지>



(http://word.snu.ac.kr/kosac/index.php)

저희는 감성분석을 하기위해 텍스트 데이터에 극성을 달아야 했습니다.
이를 위해, 저희는 KOSAC(한국어 감정 및 의견 분석 코퍼스)를 활용하였습니다.

```
text = "사랑해요 투빅스"

txt_k = kkma.pos(text)
txt_k
```

```
[('사랑', 'NNG'),
 ('하', 'XSV'),
 ('어요', 'EFN'),
 ('투', 'NNG'),
 ('빅', 'NNG'),
 ('슬', 'VV')]
```

<형태소 분석 예시>

KOSAC 감성사전을 이용하기 위해서는 먼저 **형태소 분석**을 해야합니다.
형태소 분석기는 KOSAC 감성사전의 형식과 가장 유사한 형태로 나눠주는
KONLPY에 내장되어 있는 '꼬꼬마 형태소 분석기'를 활용하였습니다.

③ 프로젝트 과정
3) 감성분석

	A	B	C	D	E	F	G	H	I	J
1		Inten-H	expr-I	expr-da	expr-de	expr-ds	expr-wd	inten-L	inten-M	inten-N
2	0	3.796578	9.259459	0.541324	11.45394	11.45561	3.289674	3.578908	25.49008	3.134433
3	1	4.160713	8.909123	0.734143	12.56231	18.10802	5.686405	4.320725	32.06172	5.456844
4	2	2.738262	4.388276	0.011111	5.183975	8.292806	2.123831	2.163845	12.30763	2.790258
5	3	7.735428	8.45102	0.037366	8.395229	17.35531	4.761074	4.15061	22.94254	4.171419
6	4	1.585714	4.773485	0	7.497835	12.33149	2.397186	3.384336	19.67443	2.355519
7	5	0.525499	1.667704	0.006452	0.938978	2.332124	0.054743	0.748272	3.660407	0.065822
8	6	0.501595	1.122678	0.037366	3.631041	4.604915	2.604	1.184917	8.760141	1.553347
9	7	5.942323	5.0947	0.189973	12.62555	19.32692	2.762864	6.251678	24.81878	2.987223
10	9	3.550417	9.567515	0.059938	10.67204	15.41815	4.282362	5.094851	28.15245	3.202287

…

⋮

앞서 말한 과정들을 통해,
영화 평들은 총 29가지 변수를 가진 데이터가 됩니다.

프로젝트 과정
3) 감성분석

	A	B	C	D	E	F	G	H	I
1	X	review_data	스토리	재미	연기	이해	긴장감	classified	type
2	similarity	영화를 보는	1.392846	1.572395	0.61516	0.751653	1.946108	긴장감	neut
3	similarity	분위기로 보는	0.825231	1.104023	0.771678	0.694231	0.838665	재미	neg
4	similarity	뻔할 수 있는	2.481165	1.795907	4.219205	0.826535	1.626136	연기	pos
5	similarity	무섭지만은 않	1.338093	1.493557	0.857219	1.133663	1.470902	재미	neg
6	similarity	'전반부는 괜	3.297156	3.326167	4.158264	1.382304	2.665082	연기	neg
7	similarity	2번 봤는데 2	1.147155	1.970664	0.933674	1.331418	2.188829	긴장감	pos
8	similarity	배우들의 맛깔	2.736745	2.010845	4.106133	1.585055	2.246647	연기	neg
9	similarity	시간 가는 줄	3.194446	2.307892	4.08909	1.663725	3.231101	연기	neg
10	similarity	곡성의 가장	1.013364	1.328693	0.933065	0.664448	0.983112	재미	neg
11	similarity	영화에 호불호	3.008921	3.804916	4.085896	2.91862	2.870648	연기	pos
12	similarity	내가 생각했던	3.908174	3.224646	4.082342	2.7441	3.480821	연기	neut
13	similarity	집 가서 자꾸	1.363797	1.840394	1.008006	2.683451	1.405013	이해	neg

이 데이터를 분석하여 긍정/중립/부정으로 분류하였습니다.
분류기는 여러 분류기를 사용해본 결과 가장 좋았던
Random Forest(약 60%)를 사용하였습니다.

4. 결과 및 제언

평가 카테고리

〈재미〉

Positive

"호러 스릴러, 영적인 소재를 많은 영화를 좋아하는데
한국에서 이런 작품을 만들 줄이야.
긴 시간이었는데도 지루한 감이 없었네요."

"아리송하기만 한 주인공들의 관계,
재미보다는 관객을 헷갈리게 하는데 온 정신이
쏠려있는 것 같아서 아쉽지만 나름 훌륭한 작품 같습니다."

"2번 봐도 온 몸에 힘이 다 빠지네요. 개연성 따지시는 분들
이 영화가 다루는 주제 자체가 카오스합니다.
몰입시키는 힘은 최근 본 영화 중 최고네요."

Negative

"무슨 말인지는 알겠는데 그냥 재미가 없다."

"재미없었음. 진짜 기대하고 봤는데 소름 끼치고
긴장되는 분위기는 잘 조성한 것 같으나 별 감동 없었음."

"목성 해적 떠 있길래 해적 보고 이해하고 나니 재미있었음
근데 재미없었음 평은 뭐지?"

평가 카테고리

〈 이해 〉

Positive

"정말 집중해서 봤어요. 이해 안되는 부분은
다른 분들 글 쏟아 보구 이해했더랬지요.
생각할수록 뚜서워서 떠철 잠 설쳤네요."

"반전에 반전 연기력도 대단
어느 정도 이해하고 보면 재미있는 영화가 될 듯 해요"

"두 번이나 봤어요. 첫 번째 봤을 때 재밌기는 해도
내용이 잘 이해가 안 가더라고요. 두 번 봐도 갸우뚱"

Negative

"재밌었는데 내용이 완전히 이해되지 않았다."

"불만했습니다,,, 문제는 처음부터 끝까지
잘 봐야 이해가 간다는 것"

"돈 주고 스트레스받는데 볼만한 영화.
두 번은 봐야 되는 영화"

평가 카테고리

〈 연기 〉

Positive

"무서운 장면 하나 없이 어마어마하게 무섭다.
연출과 연기, 그리고 쿠키까지 모두
어린 아역배우에게 압도당해버렸다."

"스산한 분위기가 작품 배경, 이미지랑 잘 맞고
배우들의 열연이 돋보인다."

"곡성 대박 단 1초도 아까움이 없었던 영화 .
락도원 황정민의 케미 다음에도 또 이렇게 이길"

Negative

"연기자의 연기가 만점 10점 만점에 10점
근데 귀신 얘기로 흘러 들어가는 것 같아 집중이 안되고
음산한 분위기와 배우들의 연기로 몰입되는 영화지 내용은 그냥."

"배우들이 연기를 잘해서 2점이라도 줬다.
이렇게 잔인하고 무서운데 어찌 15세 관람가라는 건지.
황정민의 정체는 무엇이며 천우희의 정체는 뭐라는 건지 .
결말을 관객에게 미루는 건 뭐람,
무서울과 답답함 찝찝함만 남는 영화."

평가 카테고리

〈스토리〉

Positive

"결론 재밌습니다 영화가 담는 메시지는 진지 합니다
한국적인 해학적인 요소들도 있고 하여튼 재밌어요"

" 마지막까지 어느 쪽이 문제인지 갈등을 고조시키는 구성이
정말 좋았고 연기자들의 집중력도 대단해서
보는 내내 긴장감을 놓을 수 없었습니다"

"연기 줄거리 지루할 조차 찾아볼 수 없을
편집 수준 또한 대단하다. "

Negative

"두서 없고 억지스러운 스토리 도원이 형 캐스팅 미스인 거 같고
여하튼 오래간만에 실망작 이네요"

"이건 뭐 스토리는 발로 쓴 것보다 못하다 평점 1점도 아깝다"

"소설로 원작이 있을 것 같을 정도로 스토리 전개가 좋았습니다
갑득의 숨겨진 본성을 찾아보는 재미가 꽤 쏠쏠하더군요"

평가 카테고리

Positive 〈 긴장감 〉 Negative

Positive

"일주일 사이에 두 번이나 봄 처음 볼 때는 긴장감으로
두 번째는 스토리 100% 이해로 재미가 더 함 "

"최고입니다 왜 호불호가 갈리는지 이해가 안 갑니다 구성면에서 완벽합니다
추격자 처럼, 한시도 긴장감을 놓을 수 없는 진짜 명작"

"너무 재밌게 시간 가는 줄 모르고 봤어요 상영시간이
이렇게 긴 줄 몰랐는데 지루하단 생각이 전혀 안 들었네요"

Negative

"초반 코믹과 스릴을 넘나드는 재미가 좋았는데,
갈수록 스릴만이 계속되는 긴장감에 힘들었답니다
끝나고 나서는 알 수 없는 찝찝함이 남음"

""곱씹으면 씹을수록 단맛이 나는 쌀알 같은 영화입니다"

"전체적인 내용과 긴장감은 좋았으나 너무나 많은 복선을
심어놓은 까닭에 본연의 스토리에 집중을 할 수 없었다"

④ 결과 및 제언
2) 제언

〈맞춤영화리뷰분석기〉 완성!
사실 아쉬운 점이 많다..

④ 결과 및 제언
2) 제언

그래도 word2vec을 통해 평가 요소별
카테고리를 나누는 것은 성공적이었고,

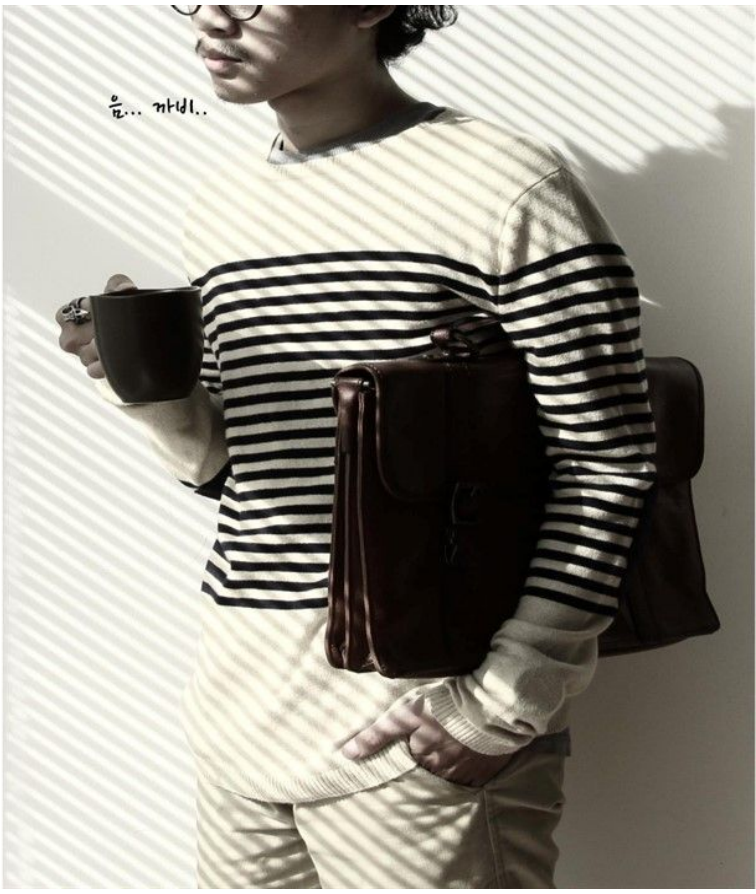
④ 결과 및 제언
2) 제언

먼저, 감성분석의 정확도가 약 60%밖에 안 되어 실사용은 힘들다는 점..

하지만, 3가지 무작위 분리한다 했을 때 정확도는 33%이며, 우리 모델의 Lift 값이 2이기 때문에 괜찮은 편이다.

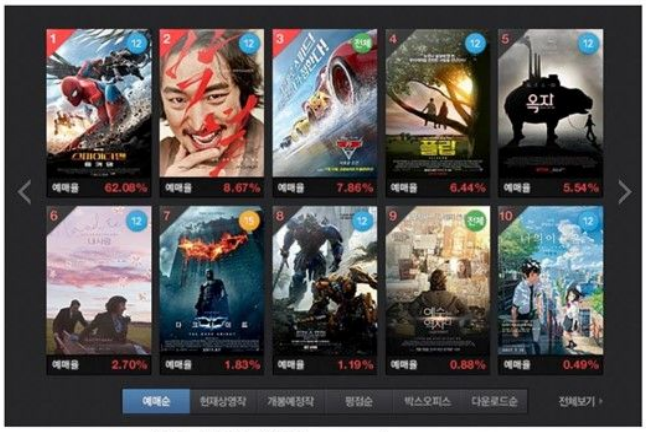
	A	B	C	D	E	F	G	H	I
1	X	review_data	스토리	재미	연기	이해	긴장감	classified	type
2	similarity	영화를 보는	1.392846	1.572395	0.61516	0.751653	1.946108	긴장감	neut
3	similarity	분위기로 보는	0.825231	1.104023	0.771678	0.694231	0.838665	재미	neg
4	similarity	편할 수 있는	2.481165	1.795907	4.219205	0.826535	1.626136	연기	pos
5	similarity	무섭지만은 않	1.338093	1.493557	0.857219	1.133663	1.470902	재미	neg
6	similarity	'전반부는 괜	3.297156	3.326167	4.158264	1.382304	2.665082	연기	neg
7	similarity	2번 봤는데 2	1.147155	1.970664	0.933674	1.331418	2.188829	긴장감	pos
8	similarity	배우들의 맛깔	2.736745	2.010845	4.106133	1.585055	2.246647	연기	neg
9	similarity	시간 가는 줄	3.194446	2.307892	4.08909	1.663725	3.231101	연기	neg
10	similarity	곡성의 가장	1.013364	1.328693	0.933065	0.664448	0.983112	재미	neg
11	similarity	영화에 호불호	3.008921	3.804916	4.085896	2.91862	2.870648	연기	pos
12	similarity	내가 생각했던	3.908174	3.224646	4.082342	2.7441	3.480821	연기	neut
13	similarity	집 가서 자꾸	1.363797	1.840394	1.008006	2.683451	1.405013	이해	neg

이 데이터를 분석하여 긍정/중립/부정으로 분류하였습니다.
분류기는 여러 분류기를 사용해본 결과 가장 좋았던
Random Forest(약 60%)를 사용하였습니다.



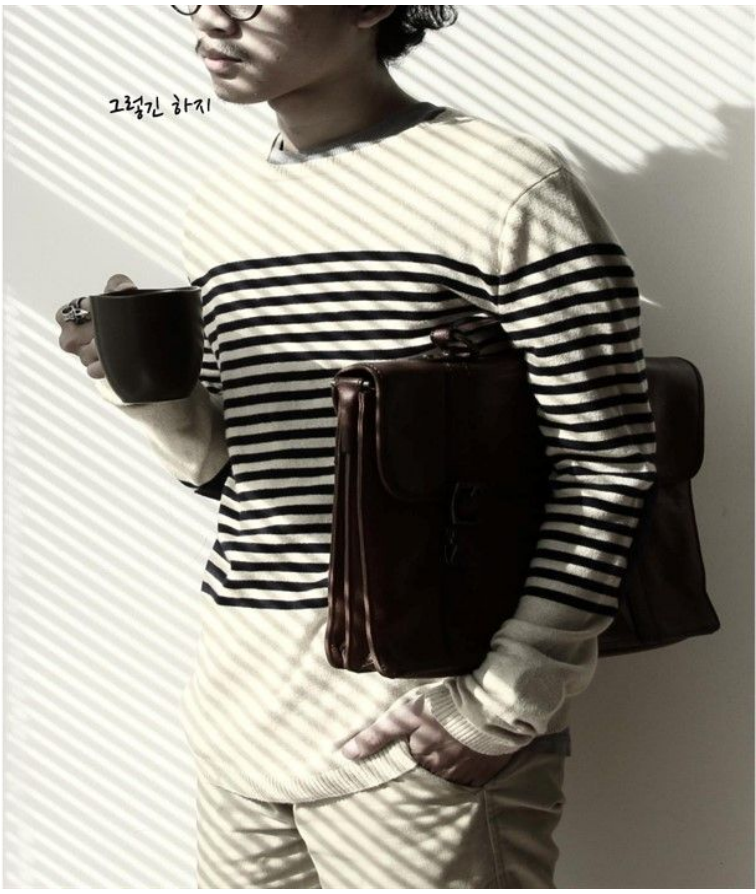
④ 결과 및 제언
2) 제언

그리고 〈곡성〉에만 적용해본 것이라는 점.



(출처 : 네이버 영화(http://movie.naver.com/))

**하지만, 이는 다른 영화에 같은 코드를 사용하면,
각 영화 별 분석이 가능해진다.**

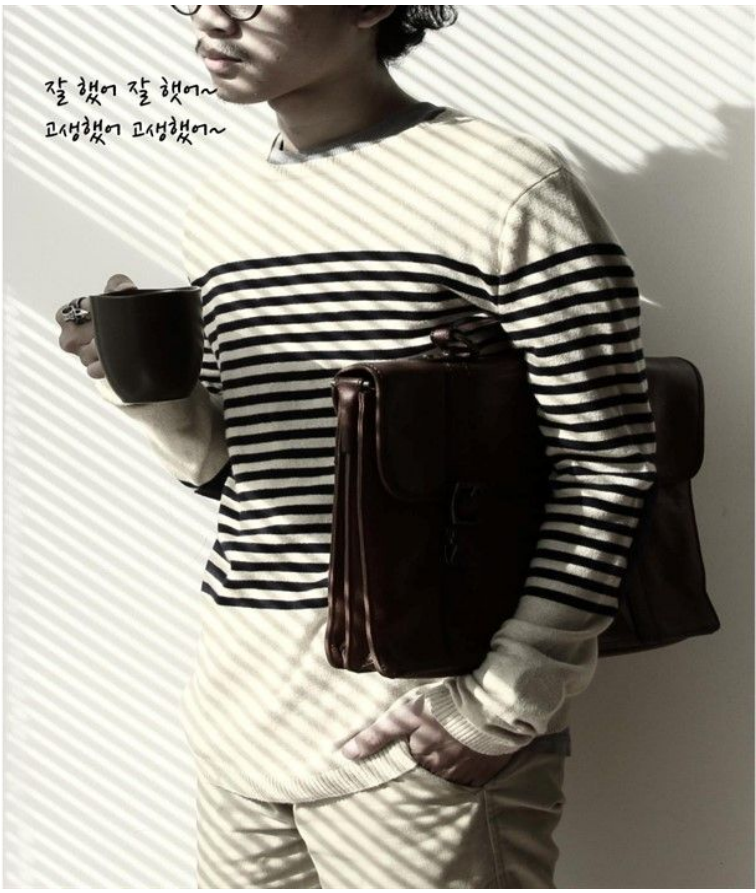


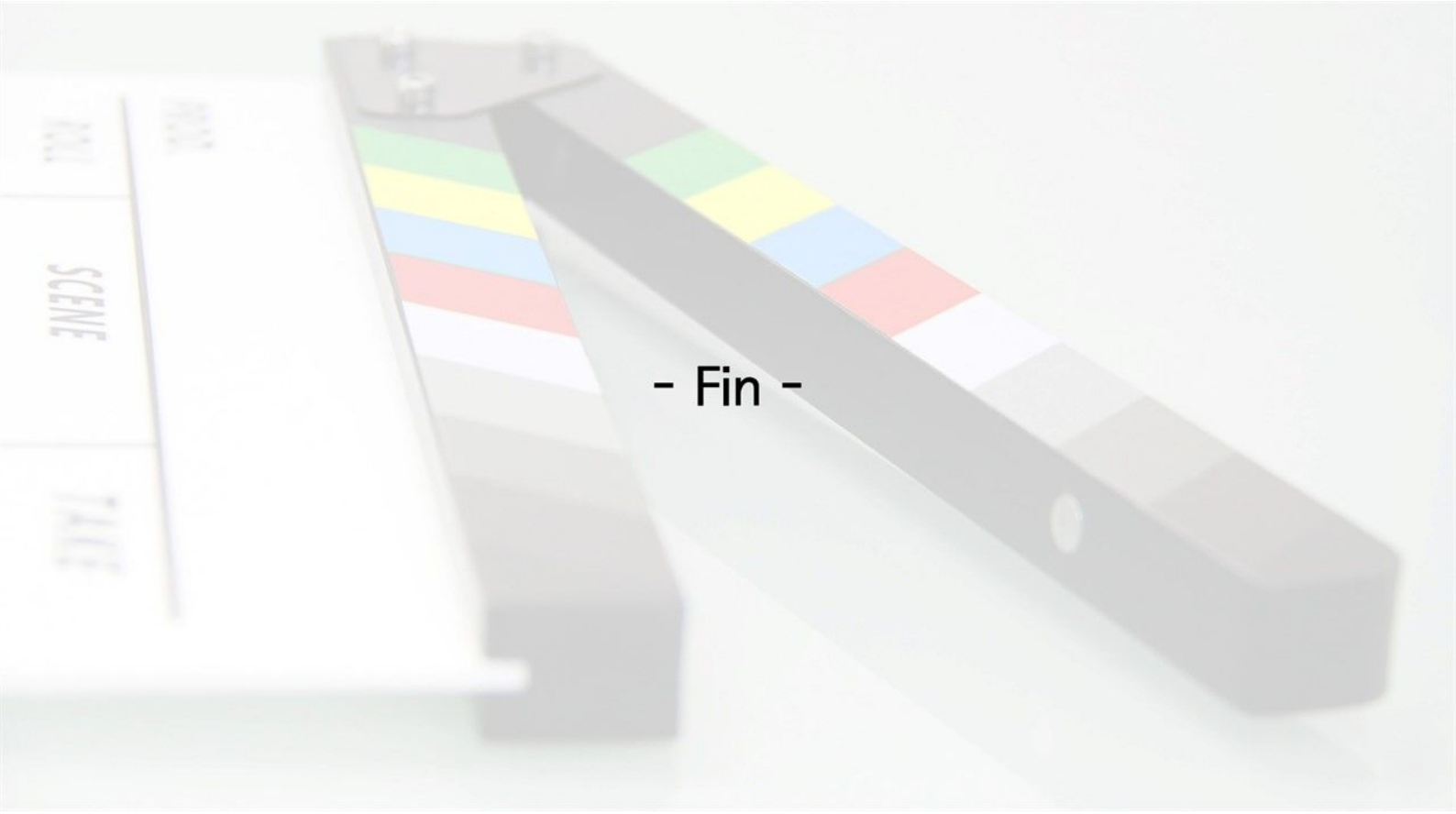
④ 결과 및 제언
2) 제언

전체적인 결과물은
정확도와는 별개로 생각보다 괜찮았고,

④ 결과 및 제언
2) 제언

신조어를 감성사전에 추가하는 등
다양한 시도를 해봤다는 점은
의미 있었다고 자체평가 해본다..ㅎㅎ





- Fin -