




Melody flows through Bamboo

Tobig's 5th PROJECT

DJ Bamboo

이제형 최도현 전종섭 김지수 이수빈



내리는 비는 우산을 두드리는 느낌이 없이 흙의 냄새를 은은하게 풍긴다

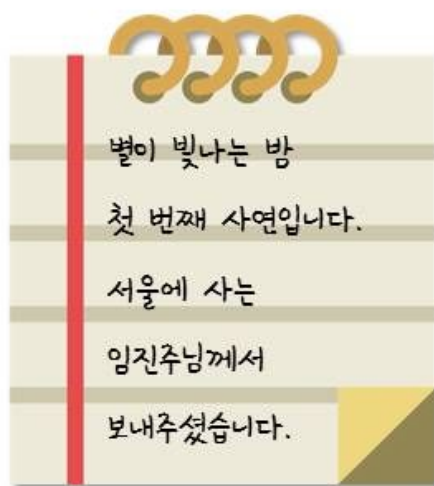
왜인지 익숙한 공기를 마시며 타야 할 버스를 기다리는 중엔
우리가 같이 타곤 했던 추억도 지나간다

실수인 척 저 버스를 타면 네게 갈 텐데

멀쩡한 정신에 차마 발을 올리진 못 해 멍하니 지나가는 모양만 지켜본다.

고려대학교 대나무숲 中
BGM 소나기 - 아이오아이

01 주제



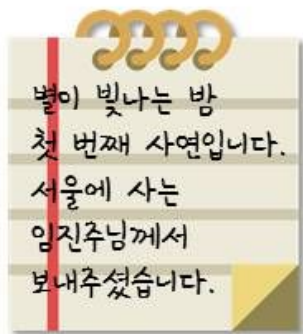
라디오에 올라오는 사연,
그리고 사연에 어울리는 노래가
주는 자그마한 감동.



INDEX

01 주제 **02** 데이터 수집 **03** 데이터 정제 **04** 데이터 분석 **05** 결과 및 제언

01 주제



+



+



대숲 사연(Text)기반
음악 추천 시스템 구현

걱정 고민 많은 20대 청춘들이
사연을 많이 올리는 '대나무숲'

진행과정

데이터 수집 ➤ 데이터 정제 ➤ 모델링 ➤ 추천시스템 구현 ➤ 결론

-고려대학교 대나무숲
-경희대학교 대나무숲
-멜론 음악 정보데이터

진행과정

데이터 수집 ➤ 데이터 정제 ➤ 모델링 ➤ 추천시스템 구현 ➤ 결론

진행과정



02 수집

1) 음악 데이터 수집

대형 음원 사이트 Melon에서 노래 정보 크롤링

NO	곡명	아티스트
1	▶ + Let's Be Happy	From Paris
2	▶ + 여름을 날다	서른의 꿈
3	▶ + A Sailor (합해)	클라인 러브
4	▶ + Daisy Day (tvN '꽃보다 ...	Tom oyuki Asakawa
5	▶ + You're So Lovely	로맨틱망고
6	▶ + 여름 가득한 푸름	물고기꿈

노래 정보 리스트
곡명, 아티스트, 가사
정보를 담고 있음

테마/장르

DJ들이 선곡한 플레이리스트를 기호와 상황에 맞게 골라서 들어보세요.

플레이리스트

플레이리스트명으로 검색해주세요

검색

테마	장르			
전체	비오는 날	여름학기	클럽	사랑/애틀
휴식/힐링	추억/회상	카페	잔잔한	여행/산책
드라이브	밤/새벽	이별/슬픔	운동할 때	기분전환
매장음악	스트레스	봄의철초	가을동화	겨울연가

테마 별로 분류 되어있는
멜론 DJ 플레이리스트.
곡의 테마 정보를 담고 있음

02 수집

1) 음악 데이터 수집

곡명, 아티스트, 가사

title	artist	lyric	랩, 힙합	휴식, 힐링	카페	비오는 날	발라드	R.B.Soul	사랑, 설렘	여행, 산책
아기공룡 하나	만화천국	아기공룡 하나	0	0	0	0	0	0	0	0
로봇 수사관	NA	침자게 달아	0	0	0	0	0	0	0	0
완송이 (Pe조권)	말썰썩러기	말썰썩러기	0	0.2	0	0	0	0	0	0
날아라 슈타인	치키치키치	치키치키치	0	0	0	0	0	0	0	0
검정고무신	NA	할아버지	0	0	0	0	0	0	0	0
명탐정 코난	아침에 눈	아침에 눈	0	0	0	0	0	0	0	0
우리의 꿈	NA	내 어린 시	0	0	0	0	0	0	0	0.125
아기공룡 하나	요리보고	요리보고	0	0	0	0	0	0	0	0
비랑의 빛	오연준	사랑들만	0	0.428571	0	0	0	0	0.214286	0.071429
요술공주	NA	너와 나의	0	0	0	0	0	0	0	0
달의 요정	NA	미안해	0	0	0	0	0	0	0	0
꾸러기 수	NA	돌기	0	0	0	0	0	0	0	0
천사소녀	NA	오늘밤에	0	0	0	0	0	0	0	0.111111
고요한밤	동방신기	고요한밤	0	0	0	0	0.375	0	0.125	0
글로리아	동방신기	지극히 높	0	0.060976	0.012195	0	0.04878	0	0.02439	0
The First	동방신기	저 돌방에	0	0	0	0	0.266667	0	0.2	0
크리스마스 러브키즈	크리스마스	크리스마스	0	0	0.2	0	0	0	0.2	0
겨울아이	빅마마	겨울에 태	0	0.047619	0.142857	0	0.238095	0	0.190476	0
소중한 사	김은정	눈아만 가	0	0	0	0	0.5	0	0	0

테마

드라이브	OST	재즈	겨울연가	봄의왈츠	여름향기	댄스	가을동화	강성	트로트
0	0.333333	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0.272727	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0.272727	0	0	0	0	0	0	0	0
0	0.25	0	0	0	0	0	0	0	0
0	0.222222	0	0	0	0	0	0	0	0
0	0	0	0.071429	0	0	0	0	0	0
0	0.375	0	0	0	0	0	0	0	0
0	0.4	0	0	0	0	0	0	0	0
0	0.25	0	0	0	0	0	0	0	0
0	0.222222	0	0	0	0	0	0	0	0
0	0	0	0.25	0	0	0.125	0	0	0
0.121951	0	0	0.012195	0	0.207317	0.085366	0	0	0
0	0	0	0.266667	0	0	0.133333	0	0	0
0	0	0	0.6	0	0	0	0	0	0
0	0	0	0.285714	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0.166667
0	0.235294	0	0.117647	0	0	0	0	0	0
0	0	0	0.321429	0	0	0.071429	0	0	0
0	0	0	0.2	0	0	0	0	0	0
0	0.083333	0	0	0	0	0.333333	0	0	0

...

30만곡의 노래 정보 크롤링
곡명, 아티스트, 가사, 멜론 DJ플레이리스트에 등록 되어있는 테마 데이터를 사용

02 수집

2) 사연 데이터 수집



4 졸업한지 좀 된 이발박진 호랑이예요. 직장생활하면서 쉬는 시간이 생기면 고대
5 여러분 1월 1일은 잘 보내고 계신가요? 고대 학우분께 제 1월 1일을 자랑하고
6 (여성분들 답해주세요!) 안녕하세요. 남자입니다. 제가 어떤여성분을 알게되어서
7 대숲 한달쯤 전에 군대에서 이별했어요. 불잡고 싶은데 불잡으면 분명 돌아올까
8 2017년 1월 1일 0시 0분. 새해가 밝았습니다. 자정에 맞춰둔 알람이 울리고, 우리
9 신년을 맞아 변장하길 바라는 대숲!! 남사친 여사친때문에 논란이 많은데요! 남사
10 당신은 항상 나이에 급급해한다. 뒤처지고 있는 것 같다고 하지만 두려워 하지
11 지겨운 방학이 돌아왔다. 기말 전후로 소개팅을 통해 만났던 너 생각에 아무것도
12 어느 때와 없는 날들이다. 평소에 타던 33번 버스는 늘상 그랬듯이 우리 집 앞 정
13 새해에는..... 전역!!!!!! 전역을 하고 싶습니다!!!!!! 대한민국 군대 다 죽구
14 올 호랑이환남덜...^^ 새해 복 많이 받으시구~^^ 하시는 모든 일들도 번창하
15 이것은 유언장이 아니다 며칠에 걸친 지리한 싸움이 일단락됐다 어디서부터 잘
16 11시 45분. '이제 술술 올라가봐야하지않나?' 곧 도시 곳곳에서 피어날 불꽃들을
17 축하한다. 여러분은 벌써 올해의 1/365을 성공적으로 낭비했다. 대학생 (0/1/0)
18 즐거운 마음으로 장식한 새터(혹은 개강파티)... 나는 1X학번인데. 편하게 대해
19 몇년째 엄청 친한 여사친인데 한때 짝사랑의 대상이었지만 마음을 깔끔히 접었
20 어째서 교수님께서는 아직까지 성적을 공개하지 않으시는 걸까. 우리를 위한 나
21 고대 대나무숲 지기를 낙담을 -남으로 통일했는데... 고양이인거 ㅇㅈ하는 부
22 노트북으로 드라마 보는데 화면 어두워질 때 비치는 내 모습이 존맛이라 볼 때
23 가끔 절 이유 없이 싫어하는 사람들이 있어요. 그러면 저는 그 사람들에게 이 악
24 정유년은 붉은 달의 해입니다. 그달입니다. 양념치킨의 해라는 거죠.
25 남들 솔직히 비둘보다 그냥 예이가 더 기분 나쁘지않아요? 아 이번 학기 망했어
26 어둠은 빛을 이길 수 없다.. 2017년 새해가 밝았습니다. 어둠은 조그마한 빛만
27 글을 잘 쓰고 싶다. 이과생인 내가 이런 재능을 탐하는 건 도통놈 심보라는 걸 알
28 나는, 간질입니다 여러분이 생각하는 그 간질은 아니예요 막 거품물고 쓰러져서
29 제가 군대 가는 것 때문에 여자친구가 너무너무너무너무 많은 걱정을 해서 고만

고려대학교 대나무숲과 경희대학교 대나무숲에서 10000개의 사연 크롤링

03 정제

1) 음악 데이터 정제

테마는
그대로 사용

제거한 항목
클래식, POP, J-POP,
일렉트로니카, 월드뮤직,
뉴에이지
(이외의 외국어 노래는
수작업으로 제거)

테마/장르

DJ들이 선곡한 플레이리스트를 기분과 상황에 맞게 골라서 들어보세요.

플레이리스트

플레이리스트명으로 검색해주세요

검색

테마	장르			
전체	비오는 날	여름학기	클럽	사랑/애틀
휴식/힐링	추억/회상	카페	잔잔한	여행/산책
드라이브	밤/새벽	이별/슬픔	운동할 때	기분전환
머강음악	스트레스	봄의철조	가을동화	겨울연가

테마/장르

DJ들이 선곡한 플레이리스트를 기분과 상황에 맞게 골라서 들어보세요.

플레이리스트

플레이리스트명으로 검색해주세요

검색

테마	장르			
전체	발라드	재즈	인디&포크	R&B/Soul
클래식	POP	록/메탈	일렉트로니카	랩/힙합
OST	댄스	J-POP	월드뮤직	뉴에이지
트로트	CCM			

03 정제

1) 음악 데이터 정제

테마는
그대로 사용

제거한 항목
클래식, POP, J-POP,
일렉트로니카, 월드뮤직,
뉴에이지
(이외의 외국어 노래는
수작업으로 제거)

테마/장르

DJ들이 선곡한 플레이리스트를 기분과 상황에 맞게 골라서 들어보세요.

플레이리스트

플레이리스트명으로 검색해주세요

검색

테마	장르			
전체	비오는 날	여름학기	클럽	사랑/애틀
휴식/힐링	추억/회상	카페	잔잔한	여행/산책
드라이브	밤/새벽	이별/슬픔	운동할 때	기분전환
머강음악	스트레스	봄의살초	가을동화	겨울연가

테마/장르

DJ들이 선곡한 플레이리스트를 기분과 상황에 맞게 골라서 들어보세요.

플레이리스트

플레이리스트명으로 검색해주세요

검색

테마	장르			
전체	발라드	재즈	인디&포크	R&B/Soul
		록/메탈		랩/힙합
OST	댄스			
트로트	CCM			

03 정제

1) 음악 데이터 정제

테마는
그대로 사용

제거한 항목
클래식, POP, J-POP,
일렉트로니카, 월드뮤직,
뉴에이지
(이외의 외국어 노래는
수작업으로 제거)

테마/장르

DJ들이 선곡한 플레이리스트를 기본과 상황에 맞게 골라서 들어보세요.

플레이리스트 ▼ 플레이리스트명으로 검색해주세요 검색

테마	장르			
전체	비오는 날	여름학기	클럽	사랑/애틀
휴식/힐링	추억/회상	카페	잔잔한	여행/산책
드라이브	밤/새벽	이별/슬픔	운동할 때	기분전환
여가음악	스트레스	봄의향기	가을동화	겨울연가

30만곡 -> 3만곡으로 줄어듦

테마/장르

DJ들이 선곡한 플레이리스트를 기본과 상황에 맞게 골라서 들어보세요.

플레이리스트 ▼ 플레이리스트명으로 검색해주세요 검색

테마	장르			
전체	발라드	재즈	인디&포크	R&B/Soul
		록/메탈		랩/힙합
OST	댄스			
트로트	CCM			

03 정제

1) 음악 데이터 정제

songid	title	artist	date	genre	like	lyric	cl	랩,힙합	휴식,힐링	가페	비오는날	발라드
3753304	벚꽃 엔딩	버스커 버	2012.03.29	Rock	269439	그대여 그대	655	0.009728	0.093385	0.015564	0.003891	0.13035
4543502	아생화	박효신	2014.03.28	Ballad	222238	하얏게 피어	483	0.004902	0.102941	0.034314	0.017974	0.124183
9631530	내가 저지	임창정	2016.09.06	Ballad	182141	떠나거든 나	496	0	0.045714	0.028571	0.017143	0.062857
5719286	뱅뱅뱅 (B/B)	BIGBANG	2015.06.01	Rap/Hip h	175053	난 깨어나	1134	0.039548	0.039548	0	0.002825	0.00565
7861392	또 다시 사	임창정	2015.09.22	Ballad	170768	이름이 좀	556	0.002732	0.068306	0.013661	0.013661	0.10929
5398990	이 소설의	한동근	2014.09.30	Ballad	169433	세계가 반	558	0	0.069054	0.028133	0.012788	0.056266
3910821	바람기억	나얼	2012.09.20	R&B/Soc	167457	바람 불어	321	0.00367	0.06422	0.025688	0.009174	0.20367
9620469	우주를 줄	불빨간사슴	2016.08.29	Folk/Blues	166602	너를 너	1093	0	0.121951	0.077236	0.012195	0.00813
8158031	CHEER UP	TWICE (트	2016.04.25	Dance	164052	마일 올리	1204	0.003012	0.027108	0.003012	0.006024	0
5682700	LOSER	BIGBANG	2015.05.01	Rap/Hip h	153634	LOSER 외	918	0.104803	0.087336	0.004367	0.026201	0.008734
5389475	양화대교	Zion.T	2014.09.22	R&B/Soc	150683	우리 집에	552	0.112	0.1328	0.0112	0.0224	0.0192
4369827	금요일에	아이유	2013.12.20	Ballad	149112	화요일엔	558	0.003185	0.103503	0.041401	0.003185	0.08121
5704397	이럴거면	백아연	2015.05.20	Ballad	145392	금해서	718	0.011494	0.103448	0.027586	0.016092	0.05977
4433490	썸 (Feat. 톨소유, 정기		2014.02.07	R&B/Soc	142608	썸 (Feat. 톨	849	0.029231	0.076923	0.026154	0.001538	0.058462
3853978	강남스타일	싸이	2012.07.15	Rap/Hip h	140754	오빤 강남	781	0.023585	0.033019	0	0	0.033019
4671151	눈, 코, 입	태양	2014.06.03	R&B/Soc	138421	만안해 미	815	0.047486	0.092179	0.005587	0.013966	0.050279
3968836	되돌리다	이승기	2012.11.22	Ballad	138119	울 수 없는	596	0.002967	0.068249	0.020772	0.011869	0.198813

좋아요 수를 기준으로 약 10000개의 곡을 선별함
(인지도 있는 노래추천을 위함)

1) 음악 데이터 정제

교정 대상 문서	교정 내용
<p>LOSER 외롭이 샌 척하는 겁쟁이 못된 양아치 가을 속에 난 JUST A LOSER 외롭이 상처받은 외롭이 상처받은 머저리 더러운 쓰레기 가을 속에 난 I'M A 슬쩍히 세상과 난 어울린 적 있어 홀로였던 내겐 사랑 따윈 말뚝 있어 진지 오래 저 시간 속에 더는 네나 만나 같은 더 이상은 못 들었어 희망찬 사랑 노래 떠나나니 그저 곁들여진 대로 각본 속 에 놓아나는 숨은 <u>피애로</u> 외박였어, I'M 돌아갈래? <u>피애로</u> 난 알리 외박였어 I'M COMING HOME 이제 다시 돌아갈래? <u>하늘보다</u> 돌아갈래 어쩔 적 제자리로 언제부턴가 난 Yeah 하늘 보다 땅을 더 바라보게 했지만, 그 <u>알리</u> I'M 돼 숨쉬기조차 힘겨워진 숨 <u>한지민</u> 그 누구도 날 잡아 주질 않네 I'M A LOSER 외롭이 샌 척하는 겁쟁이 못된 양아치 가을 속에 난 JUST A LOSER 외롭이 상처받은 머저리 더러운 쓰레기 가을 속에 난 I'M A 반쪽되는 여자를 과의 내 실수 하듯만을 사랑하고 해 뜨면 <u>심중</u> <u>해인</u>이지 못 할 나의 이기적인 <u>해나</u> 때문에 망가지 버린 기쁨 <u>해나</u> 밤에 모든 것이 망가지버린 지금 멈춘 초콜릿 너의 위험한 침주 <u>있네</u> I'M 이젠 아무런 감흥도 재미도 없는 기분 나 버릴 끝에 혼자 <u>있네</u> I'M GOING 돌아갈래? HOME 나 다시 돌아갈래 예전의 제자리로 언제부턴가 난 Yeah 사람들의 시 <u>웃어보지마</u>, 그 선을 두리위만 해 주는 것조차 지겨워 <u>웃어보지마</u> 그 아무도 날 알아주질 <u>있네</u> I'M</p>	<p>옛 인터넷스 사용자기</p> <p>외롭이 상처받은</p> <p>대치여 외롭이 상처받은</p> <p>도움말 공백이나 기호를 중복해서 썼거나 잘못 사용하였습니다.</p> <p>같은</p> <p>같은</p> <p>피진 대치여가 맞지 않거나, 다른 의견이 있으면 내용을 적어서 '보내기'를 눌러 주세요. 보내기</p> <p>알리 진지</p> <p>대치여 알린 지</p> <p>도움말 '알타'가 피동의 뜻을 가지고 있습니다.</p> <p>같은</p> <p>피진 대치여가 맞지 않거나, 다른 의견이 있으면 내용을 적어서 '보내기'를 눌러 주세요. 보내기</p> <p>더 이상은</p> <p>대치여 더는</p> <p>도움말 '더'는 통사 위에 얹어서 '게속하여', '거듭하여'나 '그 위에 보태' 처럼 쓰는 부사인데, 요즘은 어떤 점으로부터 위를 뜻하는 '이상'을 축적</p>

03 정제

2) 사연 데이터 정제

R에서 KoNLP함수 이용하여
유니코드
문장부호
숫자 } 제거

```
> sayon
```

```
[1] " 제가 최근 좋아하는 남자가 생겼어요. 이 남자가 저 한테 관심이 있는지 없는지 정말 궁금해요. 저는 썸이라고 생각하는데 썸이 아닐수도 있는거잖아요 π 1년 넘게 알고 지내던 사이이고 개인적으로 단둘이 만난적은 없구 카톡만 주고 받아요.. <U+2764> 남사친과 썸이 차이는 뭔가요?<U+2764> 카톡에서 어떤 종류의 말을 주고 받으면 썸. . 인가요? 여러분.. 알려주세요<U+653C><U+3E64>慎<U+3E30><U+623C><U+3E64><U+653C><U+3E64><U+623C><U+3E38>慎<U+3E64><U+653C><U+3E64>慎<U+3E30><U+623C><U+3E64><U+653C><U+3E64><U+623C><U+3E38>慎<U+3E64>"
```

```
> removefun(sayon)
```

```
[1] " 제가 최근 좋아하는 남자가 생겼어요 이 남자가 저 한테 관심이 있는지 없는지 정말 궁금해요 저는 썸이라고 생각하는데 썸이 아닐수도 있는거잖아요 π 년 넘게 알고 지내던 사이이고 개인적으로 단둘이 만난적은 없구 카톡만 주고 받아요 "
```

03 정제

2) 사연 데이터 정제

국립국어원
맞춤법검사기로
사연데이터맞춤법 확인
Python Selenium 이용

교정 대상 문서

[총 559지]

교정 내용

옛 인터페이스 사용하기

그대는 어떻게 그렇게 예뻐요? 그대의 맑은 눈, 귀여운 입술, **오독한** 코 모두 **별이 돼요** 다정하게, 사랑스럽게 **별이 돼요**. 그 어떤 것보다 따스하게, 다정하게, 사랑스럽게 빛나는 별. 어떡하죠 별. 어떡하죠, 눈을 강아지 그대란 별이 내 앞에 밝게 빛나고, 눈을 떠도 당신이라는 세상이 내 안에 가득 들어오는대. 평범한 일상 속 당신은 자꾸만 제 세 **설현으로** 어떡하죠. 저, 아마도 상을 받게 바퀴 **설현**으로 다가오네요. 진짜 어떡하죠 저. 아마도 당신이 좋은 것 같아요. 아니 좋아요. 사실 좋다는 말, 서약한다는 말보다 제 마음을 온전히 전하고 싶은 말이 필요한데, 아직 찾지 못했네요. 내 안의 그대는 이렇게 규칙적이고, 딱딱한 확으로 규정하고, 표현할 수 없는, 나의 세상인데. 그러니까 세상인데. 그러니까 우리 같이 앞으로 잘 지내요. 그대를 위해서라면 모든 다 할게요. 타협하려면 타협하고, 당신을 업고 63빌딩을 오르라고 하면 오를 **힘들어 봐요** 게요. 우리 같이 울고, 웃고, 지치고, **힘들어 봐요**, 나의 세상을 그대와 같이 그럴게요. 그 대신 부디 제 안의 당신을 표현할 수 있을 때까지 저에게 당신의 물 **주워주세요** 감을 주워주세요. 당신의 물결으로 이 세상 모든 것을 그려볼게요. 나의 세상, 나의 그대

한 번에 300여줄씩 검사합니다.

입력 내용 **오독한**
대치어 **오독한**
도움말 '오독'이 표준말입니다.
참고 **어휘** 결합정보 관련 **합성어** 결합정보
비판 대치어가 맞지 않거나, 다른 의견이 있으면 내용을 적어서 '보내기'를 눌러 주세요. **보내기**

입력 내용 **별이 돼요**
대치어 **별이 돼요**
도움말 '돼'와 '되'는 발음이 유사하여 철자상의 오류를 범하기 쉽습니다. 어떤 진행의 상태를 나타내는 자동사는 '되다'가 좋습니다. '되어'의 준말은 '돼'로 표기해야 합니다.
(예) 그리 돼다 (X) -> 그리 되다 (O)
그리 돼지 (X) -> 그리 되지 (O)
그리 돼지는 (X) -> 그리 되지는 (O)
일이 잘 되 가다 (X) -> 일이 잘 돼 가다 (O)
돼니 (X) -> 되니 (O)
되야만 (X) -> 돼야만 / 되어야만 (O)
선생님이 되서 (x) -> 선생님이 돼서/되어서(o)

04 모델링

1) LDA

```
corpuswork <- function(x){
  cp <- lexicalize(x)
  cp$vocab <- removePunctuation(cp$vocab)
  cp$vocab <- gsub("c", "", cp$vocab)
  cp$vocab <- gsub(" ", "", cp$vocab)
  cp$vocab <- gsub(" ", "", cp$vocab)
  cp$vocab <- gsub("harater0", "", cp$vocab)
```

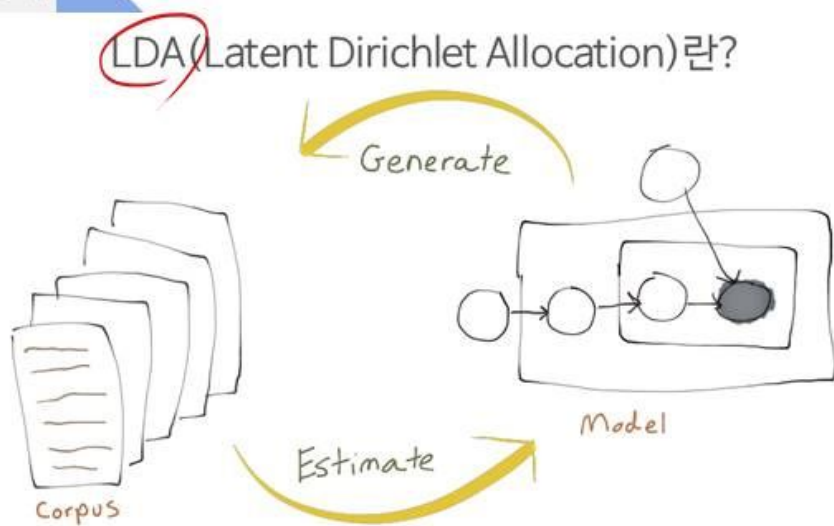
```
result <- lda.collapsed.gibbs.sampler(cor$documents, K,
  cor$vocab,
  1000, 0.1, 0.1)
```

print(top.words)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	"선배"	"엄마"	"저희"	"노래"	"오빠"	"사람"	"친구"	"사회"	"당신"	"가족사"
[2,]	"공부"	"아빠"	"제보"	"응원"	"진짜"	"생각"	"남자"	"생각"	"사랑"	""
[3,]	"학교"	"부모님"	"학우"	"손님"	"언니"	"사랑"	"여자"	"여성"	"우리"	""
[4,]	"새내기"	"동생"	"선배"	"게임"	"카테고리"	"친구"	"연애"	"동성애"	"마음"	"자리"
[5,]	"대학"	"할머니"	"행사"	"고대"	"답배"	"마음"	"애가"	"정치"	"그녀"	""
[6,]	"수업"	"아버지"	"학생"	"축제"	"다이어트"	"자신"	"연락"	"남성"	"시간"	"버스"
[7,]	"동기"	"누나"	"문제"	"어제"	"치킨"	"행복"	"자기"	"후보"	"기억"	"시간"
[8,]	"시험"	"가족"	"고려"	"티켓"	""	"모습"	"진짜"	"반대"	"순간"	"수업"
[9,]	"학기"	"어머니"	"생각"	"응원가"	"음동"	"연애"	"근대"	"사랑"	"행복"	"소리"
[10,]	"돌아리"	"할아버지"	"참여"	"연대"	"음식"	"사람들"	"사진"	"투표"	"벗꽃"	""
[11,]	"친구"	"아들"	"불면"	"진짜"	"공통화"	"이야기"	"사랑"	"참오"	"추억"	"열람실"
[12,]	"생활"	"우리"	"집단"	"만걸리"	"이거"	"하루"	"생각"	"차별"	"생각"	"화장실"
[13,]	"교수님"	"집안"	"대학교"	"입실엔티"	"문제"	"감정"	"마음"	"너희"	"하늘"	""
[14,]	"시간"	"음도"	"고대"	"가수"	"오빠"	"하계"	"군대"	"대통령"	"미소"	"오늘"
[15,]	"학생"	"전화"	"주점"	"공연"	"냉채"	"시간"	"사이"	"세상"	"그때"	"지하철"
[16,]	"생각"	"자식"	"의견"	"연예인"	"치즈"	"상처"	"커플"	"이해"	"마지막"	"카페"
[17,]	"화녀"	"저희"	"관대"	"아이"	"술로"	"사실"	"이성"	"우리"	"오늘"	"아저씨"
[18,]	"고등학교"	"논문"	"상황"	"응원단"	""	"순간"	"관계"	"존재"	"바람"	"도서관"
[19,]	"학점"	"소리"	"후배"	"아름"	""	"노력"	"관객"	"대전"	"그때"	"공부"
[20,]	"후배"	"말씀"	"단체"	"재미"	"송"	"우울"	"성격"	"문인"	"가슴"	"다리"
[21,]	"입학"	"이혼"	"대나무숲"	"고려"	"애기"	"고민"	"고백"	"나라"	"눈물"	"화장"
[22,]	"유흥"	"부모"	"발언"	"오늘"	"신발"	"누구"	"문제"	"대한"	"모습"	"교수님"
[23,]	"대학교"	"가정"	"이야기"	"영상"	"소주"	"애기"	"남자"	"민국"	"첫사랑"	"아침"
[24,]	"고대"	"강아지"	"연세"	"연세"	"모두"	"행동"	"답장"	"동성애자"	"생일"	"홍금"
[25,]	"전공"	"엄마"	"안녕"	"고대생"	"매날"	"기분"	"전환"	"선택"	"새벽"	"건물"
[26,]	"출연"	"자기"	"문위기"	"콘서트"	"관면"	"누군가"	"장난"	"의견"	"세상"	"통학"
[27,]	"안녕"	"애기"	"여러문"	"카드"	"송"	"우리"	"틀이"	"우리나라"	"겨울"	""
[28,]	"이번"	"대학"	"문화"	"만우절"	"근대"	"인생"	"하계"	"이유"	"선물"	"매일"
[29,]	"항각"	"죽음"	"사람"	"다들"	"여러문"	"세상"	"처음"	"문제"	"노래"	"어디"
[30,]	"고려"	"사촌"	"불편"	"향음"	"시험"	"그때"	"서운"	"단어"	"목소리"	"고민"

03 정제

2) 사연 데이터 정제



Unsupervised Generative Topic Model의 일종으로 문서를 모델링하는 기법
문서들을 학습하면 문서 내 단어들이 주제나 사용 용도에 따라 군집화

04 모델링

1) LDA

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

주제 이름 (Arts, Budgets, Children, Education)은 알고리즘이 자동으로 뽑아주는 것이 아니라 사람이 지정

LDA결과주제별로관련성이높은단어를뽑은 결과
-문서모델링, 키워드클러스터링에사용

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

예제문서에서주제가확실한단어에색칠됨

04 모델링

1) LDA

LDA에서 얻은 키워드 기반,
5개의 topic으로 분류
(가족, 학교생활, 사랑/연애,
사회문제, 기타)

print(top.words)

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
"선배"	"엄마"	"저희"	"노래"	"오빠"	"사람"	"친구"	"사회"	"당신"	"거숙사"
"공부"	"아빠"	"제보"	"응원"	"진짜"	"생각"	"남자"	"생각"	"사랑"	""
"학교"	"부모님"	"손님"	"연니"	"언니"	"사랑"	"여자"	"여자"	"우리"	""
"새내기"	"동생"	"선배"	"게임"	"카테고리"	"친구"	"연애"	"동성애"	"마음"	"자리"
"대학"	"할머니"	"행사"	"고대"	"답배"	"마음"	"애기"	"정치"	"그녀"	""
"수업"	"아버지"	"학생"	"축제"	"다이어트"	"자신"	"연락"	"님성"	"시간"	"버스"
"동기"	"누나"	"문제"	"어제"	"치킨"	"행복"	"자기"	"후보"	"가역"	"시간"
"시험"	"가족"	"고려"	"티켓"	""	"모습"	"진짜"	"반대"	"순간"	"수업"
"학기"	"어머니"	"생각"	"응원가"	"음동"	"연애"	"근대"	"사람"	"행복"	"소리"
"돌아리"	"할아버지"	"참여"	"연대"	"음식"	"사랑들"	"사진"	"두표"	"벗꽃"	""
"친구"	"아들"	"꽃면"	"진짜"	"공통화"	"이야기"	"사랑"	"힘으"	"추억"	"열람실"
"생활"	"우리"	"집단"	"만점리"	"이거"	"하루"	"생각"	"차별"	"생각"	"화장실"
"교수님"	"집안"	"대학교"	"양심원인"	"문제"	"감정"	"마음"	"너희"	"하늘"	""
"시간"	"음도"	"고대"	"가수"	"오빠"	"하계"	"군대"	"대통령"	"미소"	"오늘"
"학생"	"전화"	"주점"	"공연"	"냄새"	"시간"	"사이"	"세상"	"그때"	"지하철"
"생각"	"자식"	"의견"	"연예인"	"치즈"	"상처"	"커를"	"이해"	"마지막"	"카페"
"화난"	"저희"	"온대"	"아이"	"솔로"	"사실"	"이성"	"우리"	"오늘"	"아저씨"
"고등학교"	"논술"	"상황"	"응원단"	""	"순간"	"관계"	"존재"	"바람"	"도서관"
"학점"	"소리"	"후배"	"아름"	""	"노력"	"관개"	"대전"	"그때"	"공부"
"후배"	"말씀"	"단체"	"재미"	"송"	"우울"	"설계"	"존인"	"가슴"	"다리"
"입학"	"이혼"	"대나무숲"	"고려"	"애기"	"고민"	"고백"	"나라"	"눈물"	"화장"
"휴학"	"부모"	"발언"	"오늘"	"신발"	"누구"	"문제"	"대한"	"모습"	"교수님"
"대학교"	"가정"	"이야기"	"영상"	"소주"	"애기"	"남자"	"민국"	"첫사랑"	"아침"
"고대"	"강아지"	"학생회"	"연세"	"모두"	"행동"	"답장"	"동성애자"	"생일"	"홍금"
"전공"	"엄마"	"안녕"	"고대생"	"연날"	"기분"	"전화"	"선덕"	"새벽"	"건물"
"졸업"	"자기"	"문위기"	"콘서트"	"라면"	"누군가"	"장난"	"의견"	"세상"	"통학"
"안녕"	"애기"	"여러분"	"카드"	"송"	"우리"	"틀어"	"우리나라"	"겨울"	""
"이번"	"대학"	"문화"	"만우절"	"근대"	"인생"	"하계"	"이유"	"선물"	"매일"
"항력"	"죽음"	"사람"	"다들"	"여러분"	"세상"	"처음"	"문제"	"노래"	"어디"
"고려"	"사촌"	"불편"	"향음"	"시험"	"그때"	"서운"	"단어"	"목소리"	"고민"

04 모델링

1) LDA

-Sentimental Analysis

1	고양 대숲 새내기들 넘 귀여워요
1	졸업한지 좀 된 이빨빠진 호랑
0	지겨운 방학이 돌아왔다. 기말
0	11시 45분. '이제 슬슬 올라가
0	몇년째 엄청 친한 여사친인데
0	글을 잘 쓰고 싶다. 이과생인 나
1	제가 군대 가는 것 때문에 여자
0	좋아하는 마음이 너무 커서 나
1	크리스마스 트리에 얹힌 사연

Labeling

사랑/연애 노래가 전체 곡의 대다수를 차지

따라서 사랑/연애 토픽을 긍정, 부정으로
더 자세히 나눔

사랑/연애 Topic의 150개 사연을
긍정 or 부정으로 labeling 후

감성 사전을 구축해
사랑/연애 Topic을 긍정, 부정으로 **추가분류**



04 모델링

1) LDA

-Sentimental Analysis

재미	카톡	애정	어린	정성	털털	다행	쑥맥	애교	내일
0.042	0.043	0.040	0.039	0.039	0.032	0.031	0.031	0.030	0.03
시선	심정	서운	심통	원망	대다수	전역	심적	부담	문제
-0.029	-0.028	-0.026	-0.027	-0.027	-0.025	-0.02	-0.024	-0.026	-0.032

위 긍정,부정어들 바탕으로 " 사랑,연애 " 주제 2차 분류.

최종 6개의 Topic → 가족, 학교생활, 사회문제, 기타
사랑/연애(긍정),사랑/연애(부정)

04 모델링

1) LDA

-Sentimental Analysis

```
res.ridge <- cv.glmnet(x,y,family="binomial",alpha=0,
                      nfolds=4, type.measure="class")
plot(res.ridge) #lambda 값에 따라 오분류율이 달라지는 것을 볼 수 있다.
log(res.ridge$lambda.min) #최적의 lambda 값을 판단할 수 있다.(그래프에서)
plot(res.ridge$glmnet.fit,xvar="lambda")

coef.ridge <- coef(res.ridge,s="lambda.min")[,1]
pos.ridge <- coef.ridge[coef.ridge>0]
neg.ridge <- coef.ridge[coef.ridge<0]

pos.ridge <- sort(pos.ridge,decreasing=T)
neg.ridge <- sort(neg.ridge,decreasing=F)

pos.ridge[1:10]
neg.ridge[1:10]
```

단어 별 TF-IDF 값 기반으로
Ridge/ Lasso/ Elastic net 사용하여 감성어 도출

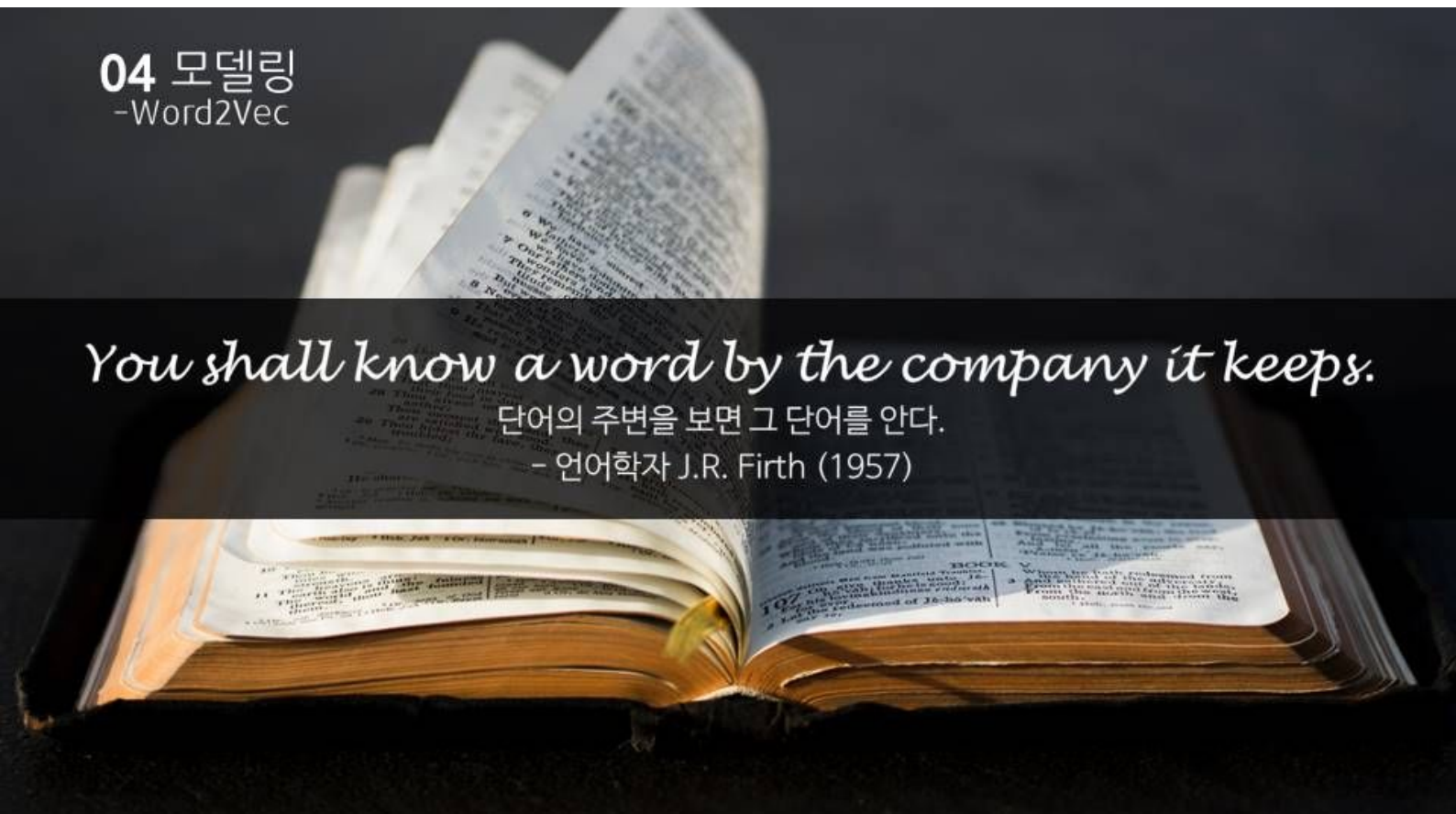
04 모델링

-Word2Vec

You shall know a word by the company it keeps.

단어의 주변을 보면 그 단어를 안다.

- 언어학자 J.R. Firth (1957)



04 모델링

2) Word2Vec

사과

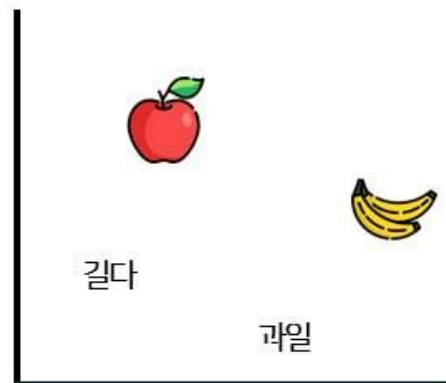


04 모델링

2) Word2Vec

사과는 맛있는 과일이다.
사과는 빨간 과일이다.
바나나는 노랑고 길다.
맛있으면 바나나, 바나나는 길어

〈학습전〉



04 모델링

2) Word2Vec

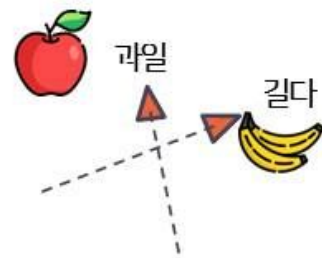
사과는 맛있는 과일이다.

사과는 빨간 과일이다.

바나나는 노랑고 길다.

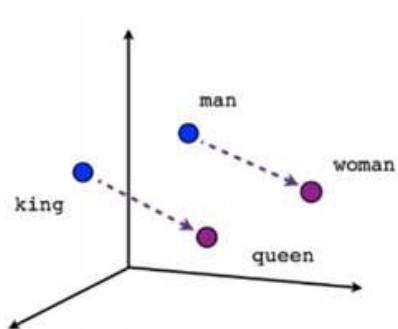
맛있으면 바나나, 바나나는 길어

〈학습 후〉

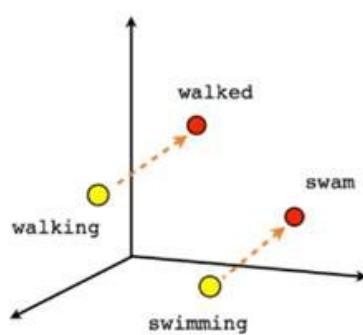


04 모델링

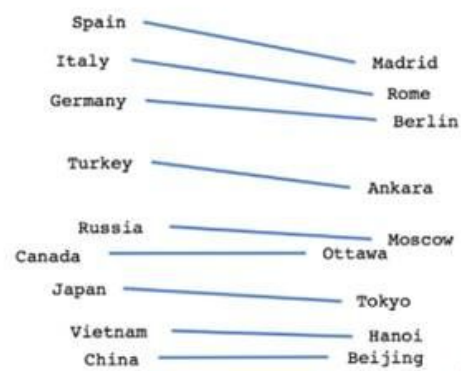
2) Word2Vec



Male-Female



Verb tense



Country-Capital

04 모델링

2) Word2Vec



학습된 벡터에서
포함된 벡터들만 추출

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
브레이크	0.359245	0.807865	-0.12907	0.798141	0.486958	0.560218	0.641586	0.122315	-0.30767	0.31942	0.22548
영자	0.585585	-0.82608	-0.00984	-0.13691	0.768528	-0.22541	0.06636	0.041538	-0.00463	-0.24042	-0.12724
스모키	-0.16996	0.323706	-0.45298	-0.24451	-0.08488	-0.63539	-0.08128	0.013616	-0.36026	0.83376	0.338799
수리	-0.30136	-0.23798	0.684479	0.629397	0.368971	-0.60992	0.870126	-0.51075	-0.48234	0.77618	0.711364
박고	-0.49459	0.107653	0.21929	0.040588	1.488544	0.519389	0.257636	-0.54444	0.246815	-0.52222	-0.47212
후회하니	-0.36013	0.017379	0.294279	0.151455	0.058024	0.710403	-0.34221	-0.73684	-0.01588	-0.0745	-0.40277
태운	-0.06833	0.187212	-0.10372	0.065602	0.392601	0.244781	-0.26287	-0.55798	0.091217	-0.07977	0.330329
까지야	-0.257	0.00787	-0.51342	0.198358	0.267835	-0.31576	0.652381	-0.33727	-0.66555	-0.78002	0.624314
합의	-0.48332	0.400826	0.357683	0.034462	0.004778	0.376516	0.041433	-0.05644	0.08197	-0.43286	0.483617
간과	-0.82924	-0.15323	0.312377	0.107098	0.104258	0.074299	0.053533	0.272262	-0.18229	0.594967	0.066793
우기	0.924772	-0.13481	0.089603	0.201542	0.810856	-0.05221	0.098439	0.28866	-0.43694	-0.1721	-0.77467
해아칠	0.591082	0.025383	-0.25033	0.006511	0.706892	0.554162	-0.15981	-0.40479	-0.33919	-0.72689	0.119037
주사	0.388895	-0.22338	-0.10391	-0.53055	-0.06696	0.279731	0.144382	-0.40585	0.536034	0.369348	1.07057
갈네	-0.57974	-0.33267	0.36219	-0.15864	-0.02787	0.35724	-0.46526	0.023413	-0.00625	0.044732	-0.12068
따라오지	-0.2726	0.371996	-0.1194	0.223755	-0.06958	-0.01515	0.557561	1.11327	-0.24347	0.119677	0.156356
재입	0.133869	0.537292	0.618846	0.441684	0.206391	0.092584	0.668783	-0.07892	0.455664	0.370112	-0.1062
바달	-0.34897	-0.10852	0.165849	0.91745	-0.01205	1.03983	0.171499	0.375307	0.683572	0.433508	1.032851
산세게	-0.07291	-0.54325	-0.05126	-0.10618	0.749004	-0.15729	0.976195	-0.20578	0.487047	0.534845	-0.39007
살금	-0.16716	0.481909	0.130344	0.156875	0.651482	0.159582	-0.22315	0.377064	0.258847	-0.90679	-0.38664



열의 평균값 = 사연 벡터

V1	V2	V3	V4	V5	V6	V7	V8	V9
0.12456607	0.29761575	-0.16206154	0.36693383	-0.05669909	-0.16114103	0.02056790	0.07072901	0.15285116
V10	V11	V12	V13	V14	V15	V16	V17	V18
0.24174693	-0.14097044	-0.62113079	0.24433360	0.30385175	0.23781069	0.06779981	0.38759459	-0.47176590

04 모델링

2) Word2Vec

'저는 잠든 엄마 품에 안기는 걸 좋아해요 엄마 품은 너무 말랑말랑하고 따뜻하거든요 참 힘든 하루를 마치고 집에 왔는데 엄마가 이미 잠들어있더라고요 평소처럼 엄마 옆에 누워서 안겼어요 조용히 엄마하고 부르니까 엄마가 졸린 목소리로 물었어요 딸 무슨 일 있어 평소와 다를 것 없이 가만히 안겼는데 속이 상해있는 걸 엄마는 어떻게 알았는지 눈물이 왁찝 터져 나왔지만 괜찮은 척 고개만 젖고 방을 나왔어요 저는 잠든 엄마 품에 안기는 걸 좋아해요 ... (생략)



명사, 동사, 형용사 추출!!

[('저', 'Noun'), ('는', 'Josa'), ('잠든', 'Verb'), ('엄마', 'Noun'), ('품', 'Noun'), ('에', 'Josa'), ('안기는', 'Verb'), ('걸', 'Noun'), ('좋아해', 'Adjective'), ('요', 'Eomi'), ('엄마', 'Noun'), ('품은', 'Verb'), ('너무', 'Noun'), ('말랑말랑', 'Adverb'), ('하고', 'Verb'), ('따뜻하', 'Adjective'), ('거든요', 'Eomi'), ('참', 'Verb'), ('힘든', 'Adjective'), ('하루', 'Noun'), ('를', 'Josa'), ('마치', 'Noun'), ('고', 'Josa'), ('집', 'Noun'), ('에', 'Josa'), ('왔', 'Verb'), ('는데', 'Eomi'), ('엄마', 'Noun'), ('가', 'Josa'), ('이미', 'Adverb'), ('잠들어', 'Verb'), ('있', 'PreEomi'), ('더라고요', 'Eomi'), ('평소', 'Noun'), ('처럼', 'Josa'), ('엄마', 'Noun'), ('옆', 'Noun'), ('에', 'Josa'), ('누워', 'Verb'), ('서', 'Eomi'), ('안겼', 'Verb'), ('어요', 'Eomi'), ('조용히', 'Adjective'), ('엄마', 'Noun'), ('하고', 'Josa'), ('부르니', 'Verb'), ('까', 'Eomi'), ('엄마', 'Noun'), ('가', 'Josa'), ('졸린', 'Verb'), ('목소리', 'Noun'), ('로', 'Josa'), ('물었', 'Verb'), ('어요', 'Eomi'), ('딸', 'Noun'), ('무슨', 'Noun'), ('일', 'Noun'), ('있어', 'Adjective'), ('평소', 'Noun'), ('와', 'Josa'), ('다를', 'Adjective'), ('것', 'Noun'), ('없이', 'Adverb'), ('가만히', 'Adverb'), ('안겼', 'Verb'), ('는데', 'Eomi'), ('속이', 'Verb'), ('상해', 'Verb'), ('있', 'PreEomi'), ('는', 'Eomi'), ('걸', 'Noun'), ('엄마', 'Noun'), ('는', 'Josa'), ('어떻게', 'Adjective'), ('게', 'Eomi'), ('알', 'Verb'), ('았', 'PreEomi'), ('는지', 'Eomi'), ('눈물', 'Noun'), ('이', 'Josa'), ('왁찝', 'Noun'), ('터져', 'Verb'), ('나왔', 'Verb'), ('지만', 'Eomi'), ('괜찮', 'Adjective'), ('은', 'Eomi'), ('척', 'Noun'), ('고개', 'Noun'), ('만', 'Josa'), ('젖고', 'Verb'), ('방', 'Noun'), ('을', 'Josa'), ('나왔', 'Verb'), ('어요', 'Eomi'), ('저', 'Noun'), ('는', 'Josa'), ('잠든', 'Verb'), ('엄마', 'Noun'), ('품', 'Noun'), ('에', 'Josa'), ('안기는', 'Verb'), ('걸', 'Noun'), ('좋아해', 'Adjective'), ('요', 'Eomi'), ('궁이', 'Noun'), ('설명하지', 'Verb'), ('않아', 'Verb'), ('도', 'Eomi'), ('모두', 'Noun'), ('위로해', 'Verb'), ('주', 'PreEomi'), ('는', 'Eomi'), ('사람', 'Noun'), ('이', 'Josa'), ('엄마', 'Noun'), ('니까', 'Josa'), ('요', 'Noun')]

04 모델링

2) Word2Vec



학습된 벡터에서 사연 벡터와 노래 벡터의 코사인 유사도 비교!
포함된 벡터들만 추출

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
브레이크	0.359245	0.807865	-0.12907	0.798141	0.486958	0.560218	0.641586	0.122315	-0.30767	0.31942	0.22548
영자	0.585585	-0.82608	-0.00984	-0.13691	0.768528	-0.22541	0.06636	0.041538	-0.00463	-0.24042	-0.12724
스모키	-0.16996	0.323706	-0.45298	-0.24451	-0.08488	-0.63539	-0.08128	0.013616	-0.36026	0.83376	0.338799
수리	-0.30136	-0.23798	0.684479	0.629397	0.368971	-0.60992	0.870126	-0.51075	-0.48234	0.77618	0.711364
박고	-0.49459	0.107653	0.21929	0.040588	1.488544	0.519389	0.257636	-0.54444	0.246815	-0.52222	-0.47212
후회하니	-0.36013	0.017379	0.294279	0.151455	0.058024	0.710403	-0.34221	-0.73684	-0.01588	-0.0745	-0.40277
태운	-0.06833	0.187212	-0.10372	0.065602	0.392601	0.244781	-0.26287	-0.55798	0.091217	-0.07977	0.330329
까지야	-0.257	0.00787	-0.51342	0.198358	0.267835	-0.31576	0.652381	-0.33727	-0.66555	-0.78002	0.624314
합의	-0.48332	0.400826	0.357683	0.034462	0.004778	0.376516	0.041433	-0.05644	0.08197	-0.43286	0.483617
간과	-0.82924	-0.15323	0.312377	0.107098	0.104258	0.074299	0.053533	0.272262	-0.18229	0.594967	0.066793
우기	0.924772	-0.13481	0.089603	0.201542	0.810856	-0.05221	0.098439	0.28866	-0.43694	-0.1721	-0.77467
헤아릴	0.591082	0.025383	-0.25033	0.006511	0.706892	0.554162	-0.15981	-0.40479	-0.33911	-0.72689	0.119037
주사	0.388895	-0.22338	-0.10391	-0.53055	-0.06696	0.279731	0.144382	-0.40585	0.536034	0.369348	1.07057
조니	0.57571	0.11191	0.35079	-0.15761	-0.12781	0.18124	0.46577	-0.552413	-0.00625	0.044732	-0.12068
재입	0.133869	0.537292	0.618846	0.441684	0.206391	0.092584	0.668783	-0.07892	0.455664	0.370112	-0.1062
바일	-0.34897	-0.10852	0.165849	0.91745	-0.01205	1.03983	0.171499	0.375307	0.683572	0.433508	1.032851
신세계	-0.07291	-0.54325	-0.05126	-0.10618	0.749004	-0.15729	0.976195	-0.20578	0.487047	0.534845	-0.39007
실금	-0.16716	0.481909	0.130344	0.156875	0.651482	0.159582	-0.22315	0.377064	0.258847	-0.90679	-0.38664



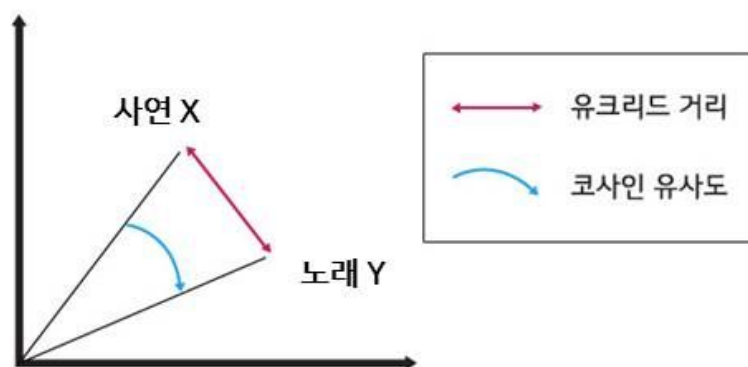
열의 평균값 = 사연 벡터

V1	V2	V3	V4	V5	V6	V7	V8	V9
0.12456607	0.29761575	-0.16206154	0.36693383	-0.05669909	-0.16114103	0.02056790	0.07072901	0.15285116
V10	V11	V12	V13	V14	V15	V16	V17	V18
0.24174693	-0.14097044	-0.62113079	0.24433360	0.30385175	0.23781069	0.06779981	0.38759459	-0.47176590

04 모델링

2) Word2Vec

코사인 유사도란 ?



05 결과

“그대는 어떻게 그렇게 예뻐요?
그대의 맑은 눈, 귀여운 입술,
오뎅한 코 모두 별이 되요.
그 어떤 것보다 따스하게,
다정하게, 사랑스럽게 빛나는 별.
어떡하죠, 눈을 감아도
그대란 별이 내 앞에 밝게 빛나고,
눈을 떠도 당신이라는 세상이
내 안에 가득 들어오는데(…)”

Play List

TITLE

ARTIST

✓ 고맙다고

원소울

✓ 나에게 그대만이

유해준

✓ TRUST YOU

원효로 1가 13-25

05 결과

“그대는 어떻게 그렇게 예뻐요? 그대의 맑은 눈, 귀여운 입술, 오뎅한 코 모두 별이 되요. 그 어떤 것보다 따스하게, 다정하게, 사랑스럽게 빛나는 별.
어떡하죠, 눈을 감아도 그대란 별이 내 앞에 밝게 빛나고, 눈을 떠도 ... (생략)”

추천 곡 1 (고맙다고-원소울)

... 늘 이렇게 내 옆에 있어줘서
정말 고맙다 고맙다 고맙다고
말할 수 없었던 내 아픔까지 함께
감싸주던 그대가 너무나 고맙다고

I pray for ya, I meditate on you all day long.

내가 살아 숨 쉬는 동안 함께 했던
많은 사람들 중 하나라면 당신이고
별종별이 하나라면 당신을 위해 빌어줄게 ...

추천 곡 2 (나에게 그대만이 - 유해준)

... 나에겐 오직 한 사람 세상
누구보다 내게 소중한 사람
널 사랑해 널 사랑해 이 말 너에게 말하고 싶어

니가 없이 난 살수 없어서
사랑 하나로만 살아가는 나에게
내 마지막 그날까지 사랑으로
남아주길 바래 그대만이 ...

05 결과

“그녀는 참 목소리를 작게 냈습니다.
단 한명만 들릴 정도로 참
작았습니다. 하지만 결코 저에게
들리지 않았던 적은 없었습니다.
조곤조곤 말을하고, 내용도 참 좋은
이야기만 하려고 두번씩 노력했던
사람이었습니다. 옷도 매일 예쁘게
입었습니다. 하지만 그때 저는
이렇게 사랑스러운 그녀에게
그만큼 잘해주지 못했습니다 (...).”

Play List

TITLE

ARTIST

- ✓ 7년간의 사랑 규현
- ✓ 우리가 정말 사랑했을까 UMC/UW
- ✓ 기억을 지워주는 병원 4 원씨겐, 팻두

05 결과

“1학기를 갓 마친 새내기예요~~
저는 2학기때 군입대를 하게됐어요...
1학기때 대학생활을 많이 참여를
못했어요 많은 친구들을 사귀지도
못하고, 동아리 활동을 많이 하지도
못하고 그렇다고 성적이 잘 나온것도
아니예요(...) 그런데 여친을 홀로
두고 군대로 가려니 너무 미안해요...
여친은 나 하나 바라보고
헌신적이어서 (...)”

Play List

TITLE

ARTIST

2학년

방탄소년단

여친 꿈을 샀다

팻두

하고 싶어

김희철&김정모

추천 시스템 구현

-추천 LOGIC

- 1 사연 입력(input)
- 2 품사 쪼개기 → 사연 분류
- 3 불용어(stopword) 제거
- 4 word2vec으로 사연의 벡터 값 도출
- 5 사연 벡터와 사연에 맞는 Topic 안의 노래 벡터와의 거리(코사인유사도)
- 6 코사인유사도가 큰 것(각이 좁은 것, 거리가 가까운 것) 3개를 output으로 추출

직접 사용해볼까요?



웹 구현 -첫 페이지



추천 시스템 구현

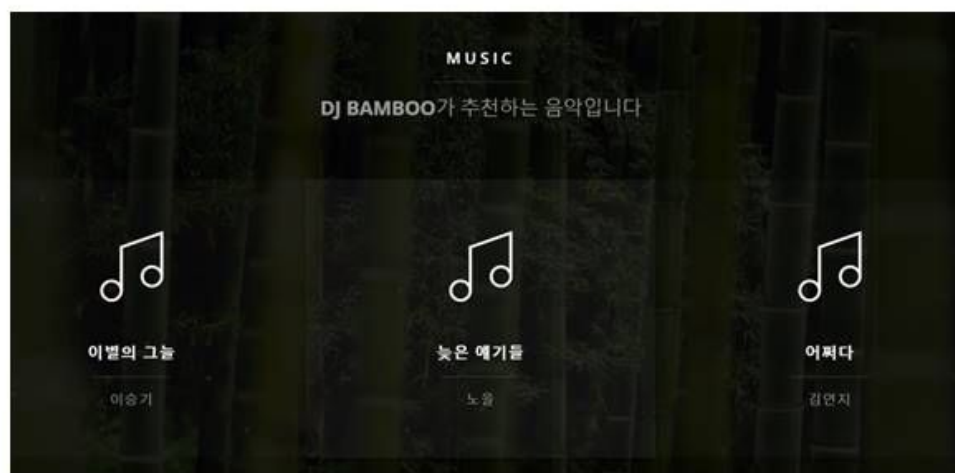
-추천 LOGIC

- 1 사연 입력(input)
- 2 품사 쪼개기 → 사연 분류
- 3 불용어(stopword) 제거, 띄어쓰기 단위 분리(Split)
- 4 word2vec으로 사연의 벡터 값 도출
- 5 사연 벡터와 사연에 맞는 Topic 안의 노래 벡터와의 거리(코사인유사도)
- 6 코사인유사도가 큰 것(각이 좁은 것, 거리가 가까운 것) 3개를 output으로 추출

웹 구현 -사연 입력



웹 구현
-노래 추천



음표 클릭 시
아래와 같은 화면으로 연결



06 제언



- 1 만족도 측정이 불가능하다.
→ Performance measure가 필요.
- 2 대학교 대나무숲이라는 특성상,
인터넷 신조어나 줄임말들의 의미도
포괄할 수 있는 더 광범위한 사전이 필요.