

일정

1단계 (4일)	<ul style="list-style-type: none">- 텍스트 형식의 뉴스 본문을 벡터로 변환시킨다.- 뉴스 간 유사도를 측정한다.
2단계 (2일)	<ul style="list-style-type: none">- 유사도를 기준으로 뉴스 본문 벡터를 2차원 벡터로 다시 변환한다.- 2차원 벡터 기준으로 유사한 뉴스 그룹을 모은다.
3단계 (3일)	<ul style="list-style-type: none">- 각 그룹별 중심이 되는 뉴스를 찾는다.- 그룹 안의 중심이 되는 뉴스와 유사하지 않은 뉴스를 찾는다.
4단계 (3일)	<ul style="list-style-type: none">- 유사하지 않은 뉴스 제거 전과 후를 비교한다.- 시각화를 위해 2차원 상 데이터 분포 시각화하여 비교한다.
4단계 (1일)	<ul style="list-style-type: none">- 최종 결과물 분석 및 중간보고서 작성
5단계	<ul style="list-style-type: none">- 1월 31일 중간 결과 보고

2023년 1월 9일 미팅 내용

- 계획서 내용 대로 2일 안에 결과 내기. -> 다음주로 연기,
- 2일안에 가능하도록, 샘플수를 줄여서 실험 결과 가지고 오기. -> 기간 연장으로 추가 실험 진행

데이터

뉴스 본문 사용

기간 : 2018-01-01 ~ 2021-04-28

카테고리 : 대기오염, 환경관리체계로 실험. (48시간 내 가능한 실험 구성을 위해 세팅)

뉴스 수

- 대기오염 : 25개
- 환경관리체계 : 25개

총 단어 수 :

- 대기오염 : 4286개
- 환경관리체계 : 4550개

차별성

- 2018~2022년 기준 약 87600개의 논문에서 이와 같은 방법을 적용한 사례 없음.

뉴스 본문 벡터 표현

1단계 : 단어의 pool과 같은 길이의 행렬을 만든다.

문서1 : 먹고 싶은 사과
문서2 : 먹고 싶은 바나나
문서3 : 길고 노란 바나나 바나나
문서4 : 저는 과일이 좋아요

-	과일이	길고	노란	먹고	바나나	사과	싶은	저는	좋아요
문서1	0	0	0	1	0	1	1	0	0
문서2	0	0	0	1	1	0	1	0	0
문서3	0	1	1	0	2	0	0	0	0
문서4	1	0	0	0	0	0	0	1	1

TF-IDF 계산법

2단계 : 행렬의 값은 각 단어의 TF-IDF 값을 넣는다.

[TF-IDF 구하는 방법]

한 뉴스에서 나온 단어 빈도 / 전체 뉴스에서 나온 단어 빈도

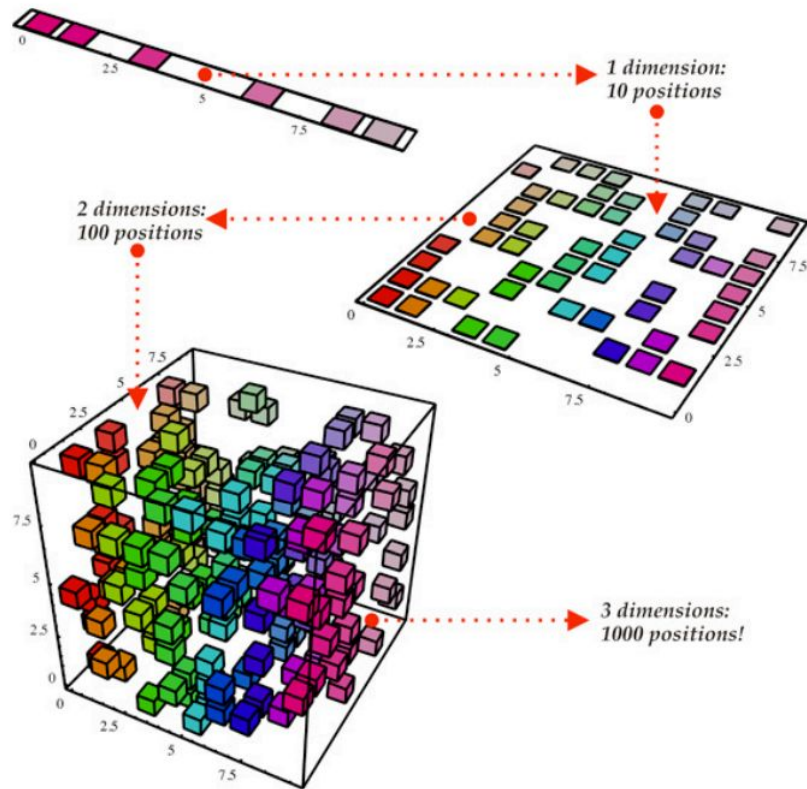
(예시) 단어 = 환경

한 뉴스에서 나온 “환경” 수 / 전체 뉴스에서 나온 “환경” 수

TF-IDF로 표현

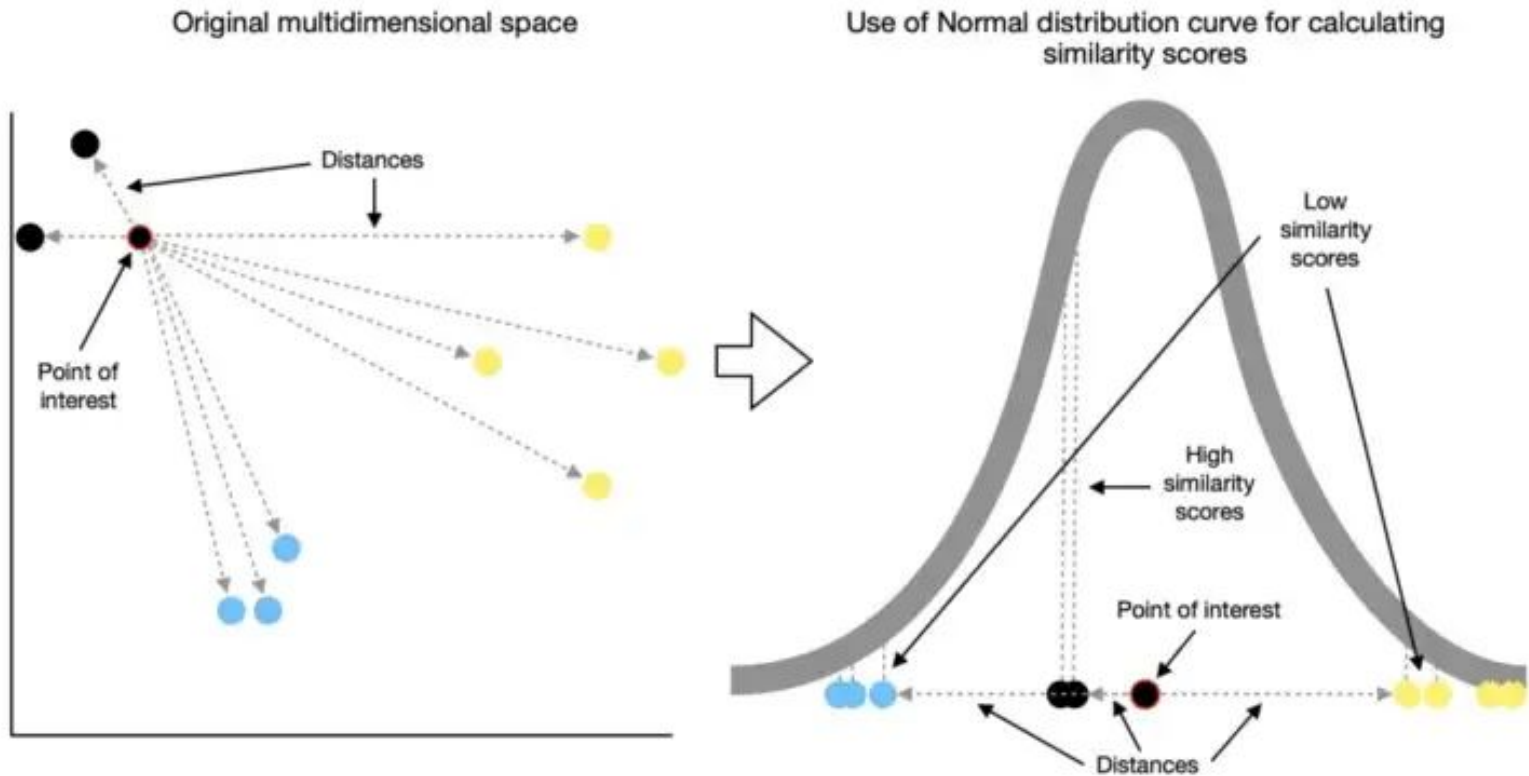
2단계 : 행렬의 값은 각 단어의 TF-IDF 값을 넣는다.

고차원 벡터를 2차원 벡터로



상자 = 뉴스 한개

고차원 벡터를 2차원 벡터로 만들기

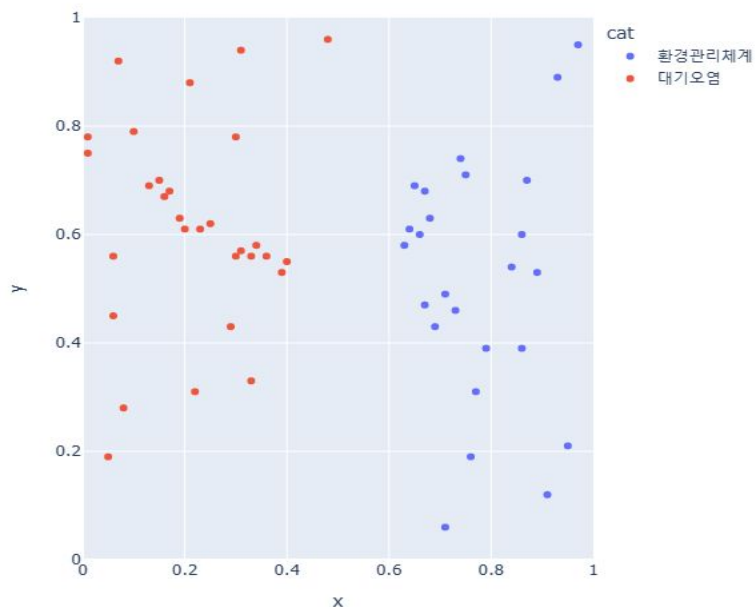


$$p_{j|i} = \frac{e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma_i^2}}}{\sum_{k \neq i} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_k||^2}{2\sigma_i^2}}}$$

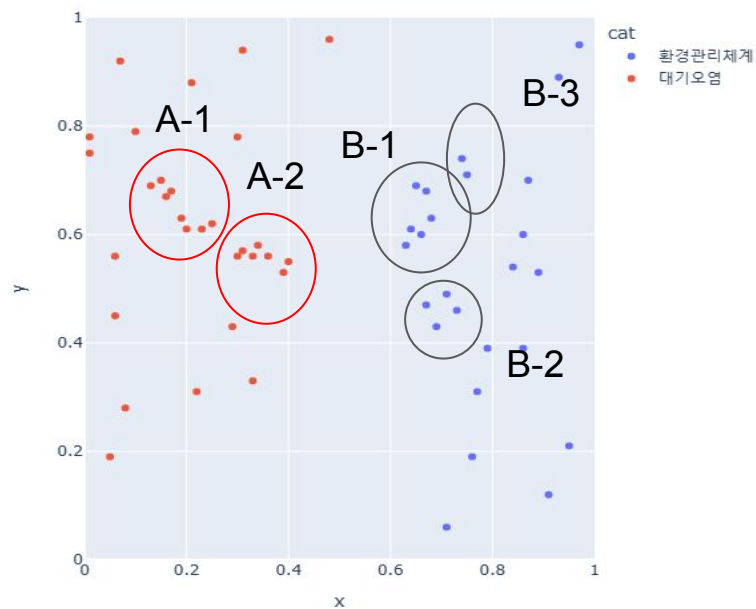
$$q_{j|i} = \frac{e^{-||\mathbf{y}_i - \mathbf{y}_j||^2}}{\sum_{k \neq i} e^{-||\mathbf{y}_i - \mathbf{y}_k||^2}}$$

본문-> 가장 작은 말의 단위로 분해 -> 벡터화 -> 2차원 공간에 시각화

분포도



분포도 : 그룹 분석



그룹 A-1 [대기오염]

공통 단어 : 오염, 배출, 대기, 물질

뉴스 예시 :

1. [‘대기오염 측정치 조작’ 영풍제련소 임원, 징역 1년 2개월 실형 | 뉴스민 \(newsmin.co.kr\)](http://newsmin.co.kr)
2. [현대제철, 방지시설 없이 오염물질 배출 - 매일경제 \(mk.co.kr\)](http://mk.co.kr)
3. [대한유화 온산공장 강력 처벌,환경오염에 경고 메시지 < 종합 < 사회 < 기사본문 - 경상일보 \(ksilbo.co.kr\)](http://ksilbo.co.kr)
4. [롯데케미칼, 대기오염물질 초과배출 부과금 취소 소송 “패소” < 경제 < 경제 < 기사본문 - 데일리시사닷컴 \(dailysisa.com\)](http://dailysisa.com)
5. ["여수산단 배출조작 기업들 사회공헌 이행의지 있나" - 노컷뉴스 \(nocutnews.co.kr\)](http://nocutnews.co.kr)
6. ['무용제 도료' 개발로 선박 도료업계 '친환경' 활기 - 아시아경제 \(asiae.co.kr\)](http://asiae.co.kr)
7. [정부, 화학물질 배출량 산정방법 개선 착수...'배출조작' 잠재울까 - 이투데이 \(etoday.co.kr\)](http://etoday.co.kr)
8. [전남도의회 "여수산단 환경개선 기업 책임 다해야" - 노컷뉴스 \(nocutnews.co.kr\)](http://nocutnews.co.kr)
9. [충남도, 대기오염물질 감축 5개 기업 친환경사업장 선정 | 연합뉴스 \(yna.co.kr\)](http://yna.co.kr)

그룹 A-2 [대기오염]

공통 단어 : 미세먼지, 산화물, 질소 U 석탄, 가스

뉴스 예시 :

1. 포스코, 친환경 청정설비로 미세먼지 감축...질소산화물 최대 80% 절감
<http://www.economytalk.kr/news/articleView.html?idxno=209194>
2. 한국가스공사, LNG공급 합작회사 매출 10년내 1조 달성
<http://mbiz.heraldcorp.com/view.php?ud=20201124000172&ntn=0>
3. 세아STX엔테크, 보령화력발전소 대기환경개선 설비 본격 가동해
<https://www.mk.co.kr/news/business/view/2020/11/1230410/>
4. 그린 일자리 최우수상 - 포스코에너지 미세먼지 종합대책 선제 대응...채용 확대 앞장
<http://mbiz.heraldcorp.com/amp/view.php?ud=20201203000425>
5. 한국지역난방공사, 대구 열병합발전소 친환경에너지 개선사업 주민설명회 개최
<http://www.newsci.com/news/articleView.html?idxno=805308>
6. 원전·석탄화력 대체한다는 LNG발전소... 전국서 잇따라 '좌초' (naver.com)
7. 현대제철, 친환경 제철소로 탈바꿈 (g-enews.com)

그룹 B -1 [환경관리체계]

공통 단어 : 뉴딜,친환경,녹색,책임,채권,영향,강화,긍정,가치,미래,감사,성과,확대,역량

뉴스 예시 :

文정부 '그린 뉴딜' 추진...은행권 친환경 사업 탄력 받나

http://m.kukinews.com/m/m_article.html?no=789382

녹색채권 발행, 재생에너지 투자... '환경'에 속도 내는 금융업계

https://news.chosun.com/site/data/html_dir/2020/08/17/2020081701479.html

민주, '그린뉴딜' 힘 보탬 녹색금융공사 설립 추진

<https://www.news1.kr/articles/?4084357>

[2020 국감] 금융사 책임투자·녹색금융 규모 **51조원**...분류체계 확립해 옥석가려야
- 매일경제

<https://www.mk.co.kr/news/economy/view/2020/10/1040243/>

NH농협금융, 환경부와 금융사 최초 **MOU** 체결

<http://www.wiseenews.com/news/articleView.html?idxno=32925>

[ESG경영] NH·삼성·KB·미래, 뉴딜 맞춰 **ESG**채권 발행 앞다퉀

https://www.fntimes.com/html/view.php?ud=20210227113047209dd55077bc2_18

그룹 B -2 [환경관리체계]

공통 단어 : 과징금, 관리, 희망퇴직, 희생, 규제, 강화, 환경
뉴스 예시 :

[2020 국감 이슈] 화력발전 5개사, 5년간 환경오염 관련 위반 94 ...

<http://www.newslock.co.kr/news/articleView.html?idxno=31817>

[2020국감] 화력발전 5사, 환경규제 위반 심각...최근 5년간 위반 94건

<http://www.straightnews.co.kr/news/articleView.html?idxno=84913>

화력발전 5사, 환경법령 위반 심각

<http://www.hkbs.co.kr/news/articleView.html?idxno=594464>

이 와중에... 르노삼성·쌍용차 300억대 과징금까지 - 조선일보

<https://www.chosun.com/economy/auto/2021/03/04/K6XVBYZ5OFHSTANRYSL3U4Z2MY/>

그룹 B -3 [환경관리체계]

공통 단어 : 농업,유통,온실가스,녹색,동물,친환경,환경친화,실천,상생,지역

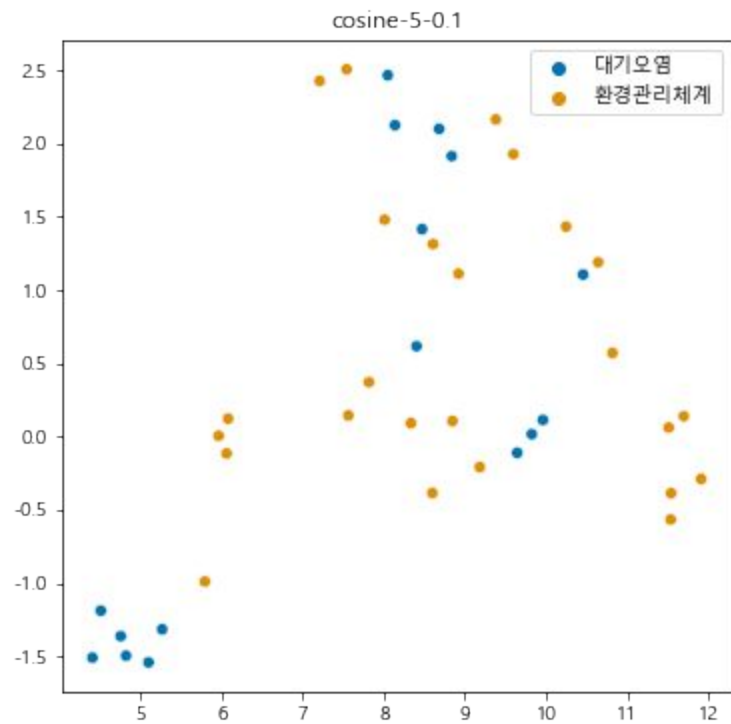
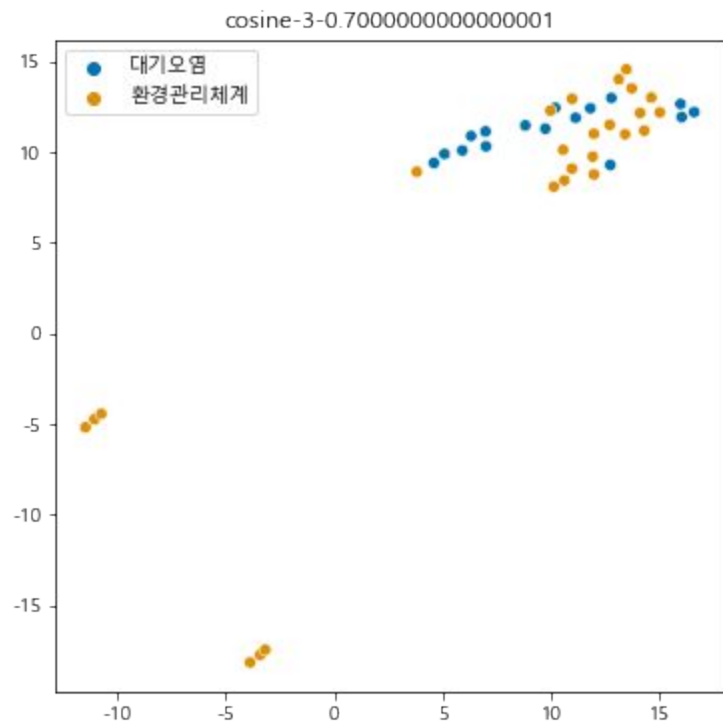
뉴스 예시 :

의식주책(衣食住冊)에 부는 '친환경' 바람

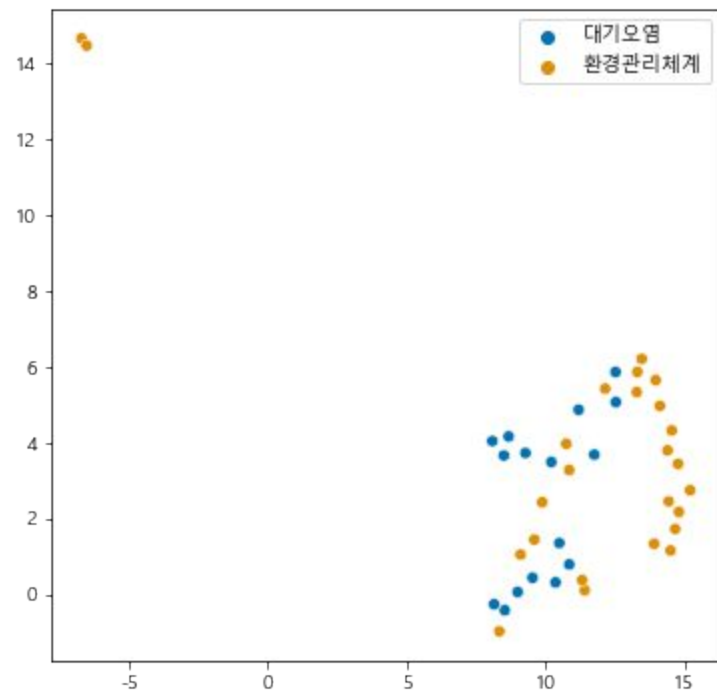
<https://www.readersnews.com/news/articleView.html?idxno=100349>

폴무원 올가홀푸드, 친환경 농업 발전 공로 '농림축산식품부 장관 ...

<https://www.koreapost.co.kr/news/articleView.html?idxno=55570>



dice-3-0.4



dice-5-0.7000000000000001

