

환경 감성사전



친환경 사업을 진행하는 기업 이미지를 위한 감성사전 구축

2021 상명대학교 데이터청년캠퍼스 3조
김진성 박채희 신한영 이건하 이아영 한현수

The background of the image shows two Starbucks mugs on a wooden table. The mug on the left is white with a green Starbucks logo, and the mug on the right is green with a white Starbucks logo. Both mugs have straws in them. The text is overlaid on the right side of the image.

GREENER STARBUCKS KOREA

단 하나뿐인 지구를 위한 약속으로 전세계 스타벅스
최초 종이 빨대 도입과 일회용 컵 줄이기 캠페인, 친환경 소재 포장재 등
친환경 활동에 앞장서겠습니다.

1 분석 배경

최근 국내 주요 대기업들은 **친환경** 키워드를 내세우며
다양한 친환경 신사업을 시도하고 있다

[ESG로 뚝다] LS그룹 "친환경 신사업으로 기업·사회 함께 성장하는 선순환 구축"

현대오일뱅크, '에코 소재·수소' 신사업 착착... HPC 황금알 겨냥

'조선대장주' 현대중, 9월 초 상장...친환경 신사업에 1조 투입

[ESG 긴급진단 ⑤] GS리테일 "친환경 제품·포장재 개발·도입 힘쓴다"

롯데정보통신, 친환경 전기차 충전 사업 시동

대기업들의 친환경 신사업

출처 네이버 뉴스 기사 제목

유튜브에서의 친환경 키워드 트렌드

2018~2019

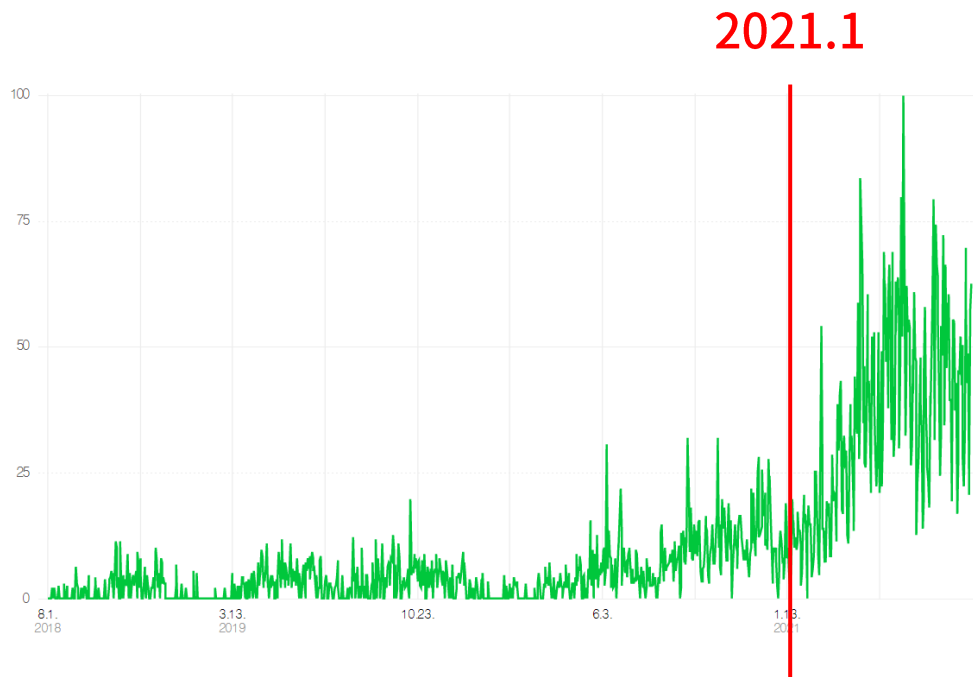


친환경에 관심이 많아지는
미디어와 소비자들

출처 Youtube 검색어

1 분석 배경

실제로 기업들의 친환경 신사업은 겉으로는 환경 친화적인 것처럼 보이지만
결국 이면에는 **환경오염**의 주범이 되고 있는 기업들의 신사업들



그린워싱 검색량

출처 네이버 데이터랩



상품에서 플라스틱의 대량 사용과
한정판이란 심리를 자극하여 자원의 낭비 유도

그에 따라 **그린워싱**에 대한 우려도 많아지는 추세

1 분석 배경

특히 친환경 사업의 대표주자인 태양광 제조와 전기차에서
환경주범이 되는 온실가스 배출을 뚜렷하게 볼 수 있다

2020 주요 원전 및 태양광제조업체 온실가스 배출량(tCO ₂ eq) 비교			
원전		태양광제조	
한수원	2,918,103	한화솔루션	2,400,795
두산중공업	229,314	오씨아이	1,258,064
합계	3,147,417	합계	3,658,859

온실가스 배출량이
원전보다 많은 태양광 사업

출처 에너지경제신문 2021.08

전체 생애주기(27km 기준, kg CO₂)



6만 1115



5만 1891

소형 내연기관차보다
탄소배출이 많은 대형 전기차

출처 한겨레 2017.12

1 분석 배경

특히 친환경 사업의 대표주자인 태양광 제조와 전기차에서
환경주범이 되는 온실가스 배출을 뚜렷하게 볼 수 있다

2020 주요 원전 및 태양광제조업체 온실가스 배출량(tCO ₂ eq) 비교			
원전		태양광제조	
한수원	2,111,031	한성에너지	1,017,795
두산중공업	229,314	오씨아이	1,258,064
합계	3,147,417	합계	3,658,859

온실가스 배출량이
원전보다 많은 태양광 사업

출처 에너지경제신문 2021.08

전체 생애주기(27km 기준, kg CO₂)

이처럼 기업들이 내세우는 **친환경 사업**은
과연 그들이 원하는 긍정적인 영향만을 불러올까?

6만 1115

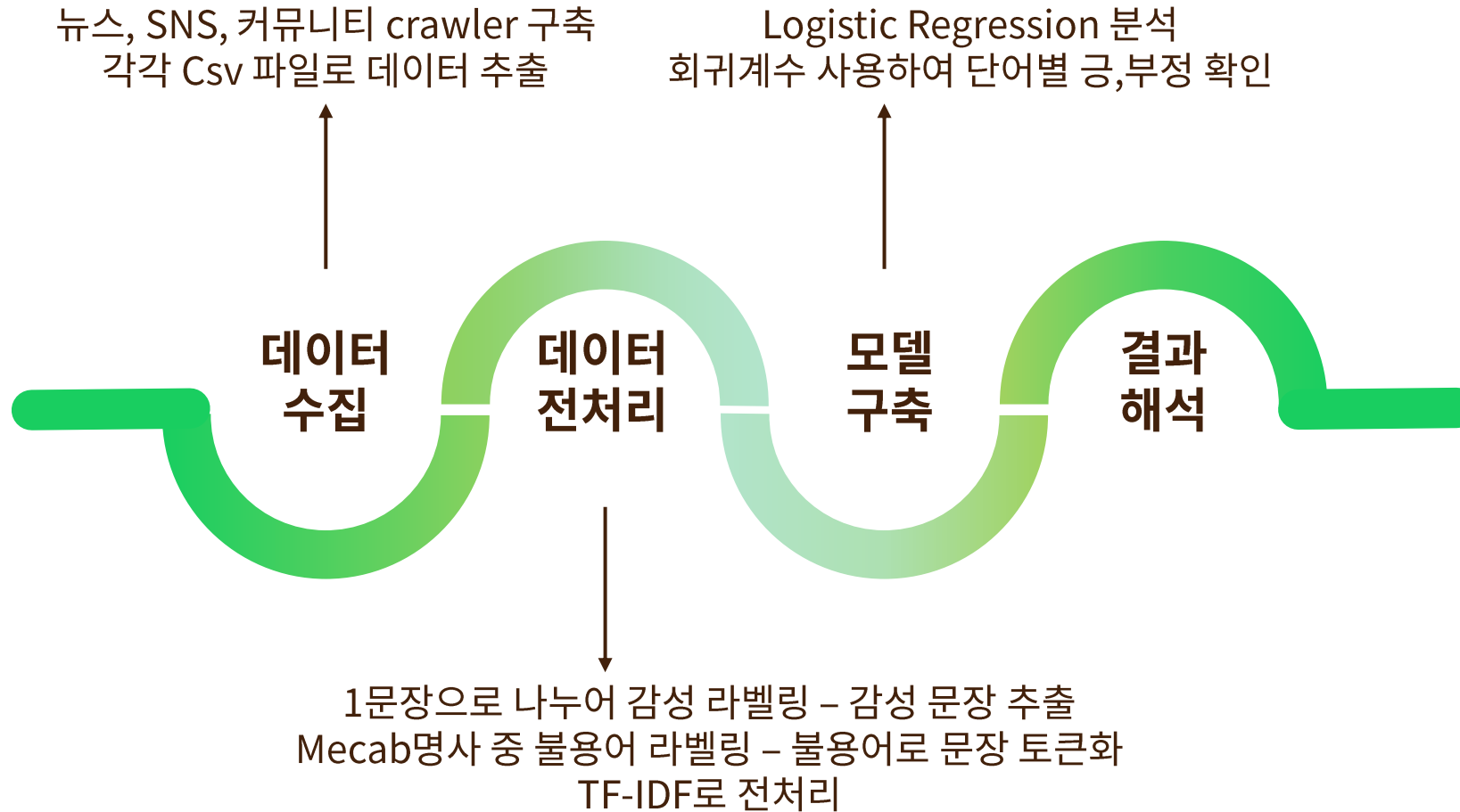


5만 1891

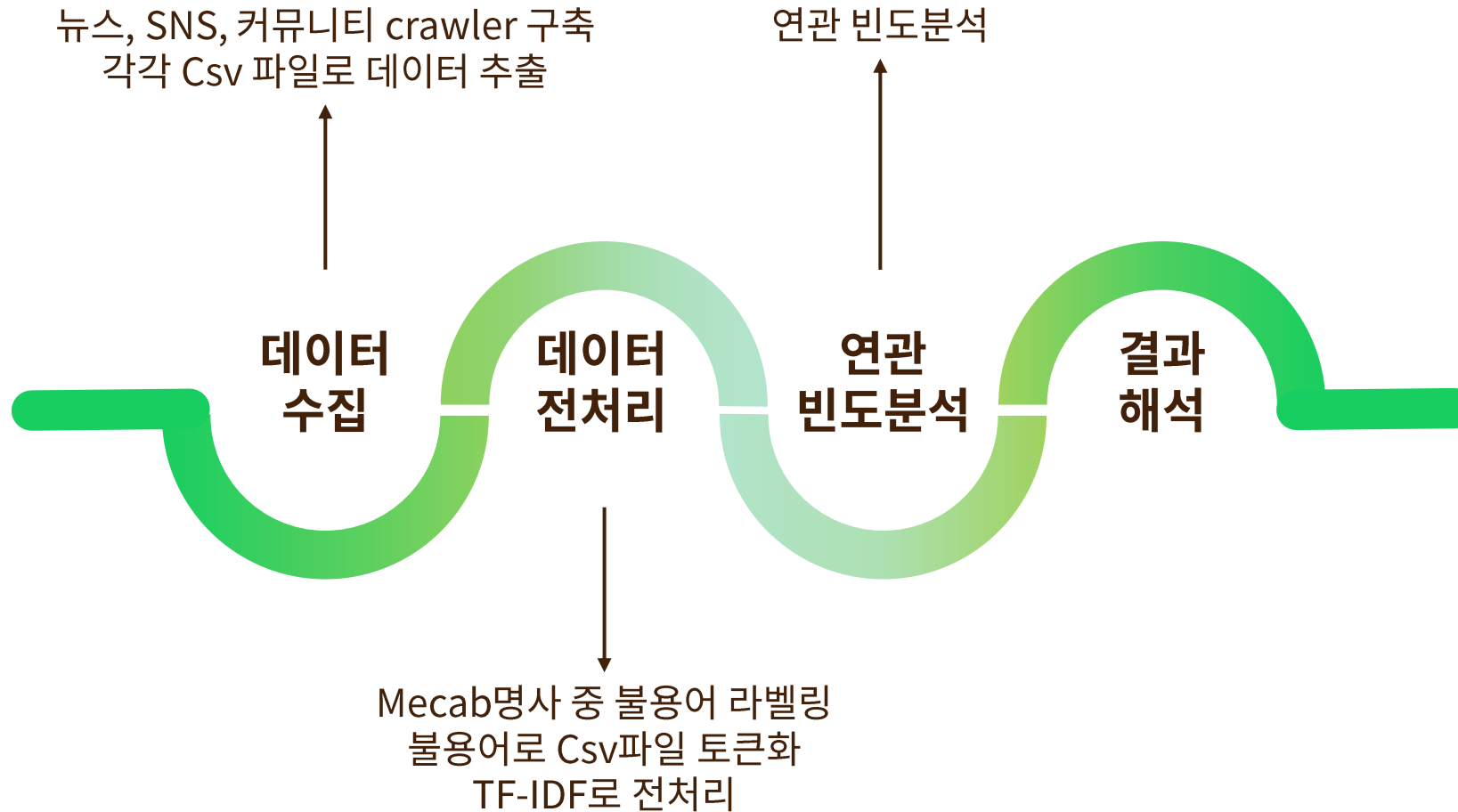
소형 내연기관차보다
탄소배출이 많은 대형 전기차

출처 한겨레 2017.12

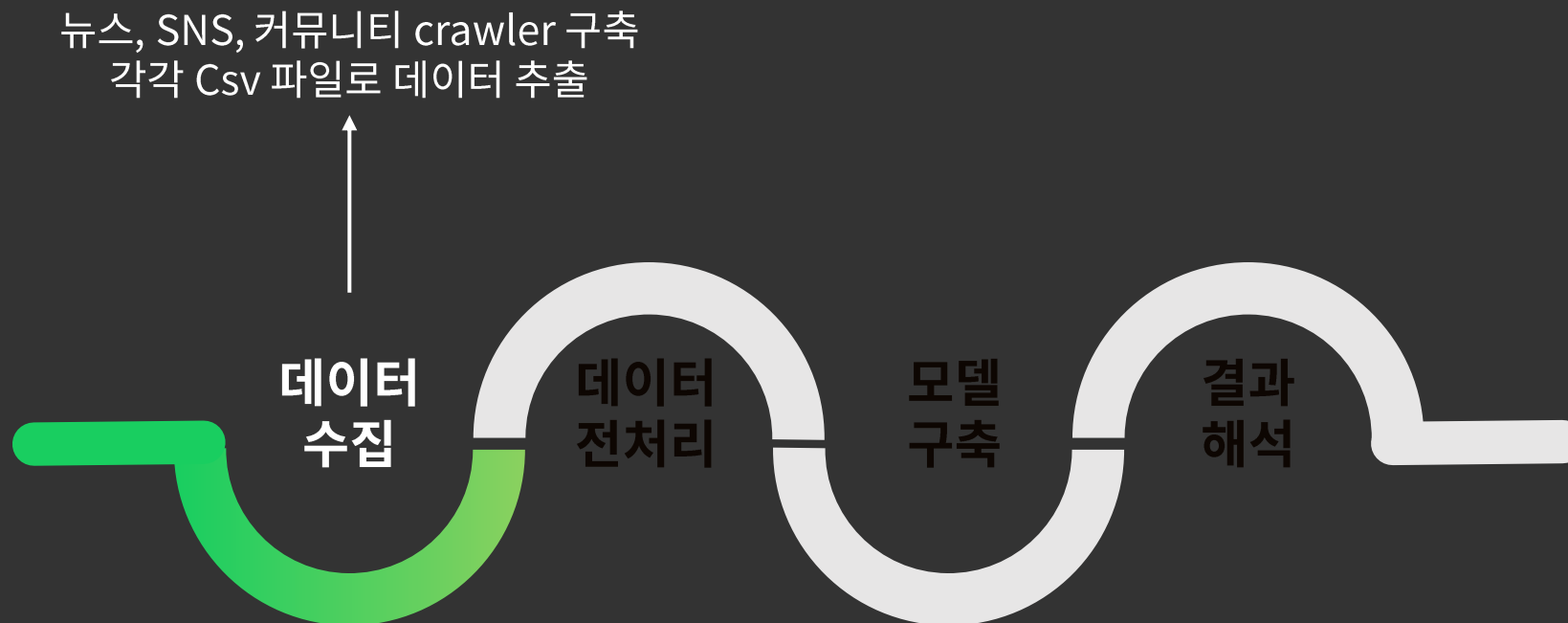
감성사전 구축 개요



연관빈도분석 개요



감성사전 구축 개요



2 감성사전 구축

데이터 수집

기업에 대한 객관적인 내용을 다룬 뉴스와
그에 대한 소비자들의 반응을 보여주는 SNS 내용 수집

#친환경 #환경오염 키워드로 검색



- 일간지 15개 언론사에서
40개씩 크롤링



- 리트윗 제외하고
게시글 300개 크롤링



- 해시태그 포함하여
키워드당 게시글 300개 크롤링

2 감성사전 구축

데이터 수집

국내 온라인 커뮤니티 월간 종합 순위 기준
상위 10개 중 크롤링이 가능한 6개 커뮤니티의 게시글 내용수집

순위	커뮤니티	투백스 지표	글수	조회수	반응수	댓글수	방문수	평균세션(초)	PPV
1	dcinside.com	64,012 -0.1683	4,163 -81	178,027,699 +9,603,838	2,285,324 +227,717	1,671,042 +162,462	158,950,000 +21,610,000	804 -16	15.97 -0.38
2	FM	58,213 +0.7719	2,151 +135	363,672,822 +26,429,227	2,207,062 +177,062	609,228 +26,118	56,300,000 +9,940,000	998 -58	25.41 -2.27
3	Q	41,835 -1.4536	12,721 +673	588,254,194 +44,815,135	1,876,612 +230,462	875,507 +26,111	24,120,000 -460,000	858 +30	12.3 +0.32
4	NATE 판	30,138 +0.8999	5,086 +155	237,175,212 +31,770,777	1,347,329 +155,075	544,444 +50,772	19,280,000 +420,000	459 +19	13.75 +0.25
5	RULIWEB	24,393 -0.157	4,017 +506	141,800,856 +7,658,241	554,747 +49,201	212,615 +13,212	43,590,000 +6,060,000	1,222 +57	20.34 -1.13
6	1	20,971 +1.0633	3,293 +875	123,380,619 +29,850,992	0 0	1,744,227 +313,985	34,010,000 +5,150,000	931 -46	14.84 -0.2
7	PPOMPPU	20,178 -0.2694	2,979 -283	85,164,180 +1,521,577	44,596 -24,270	190,589 -17,233	41,980,000 +5,540,000	666 -8	12.44 +0.32
8	INVEN	16,438 +2.8222	3,510 -221	18,046,504 +423,641	36,974 -5,223	74,162 +3,557	47,010,000 +14,180,000	557 +24	9.77 +0.34
9	CLIEB.NET	15,003 -0.1426	8,543 +358	106,447,869 +8,723,603	513,433 -71	247,295 +17,448	23,050,000 +2,530,000	685 -5	14.74 +0.11
10	인스타그램	13,637 +0.7739	824 +34	38,115,415 +1,916,393	28,408 +4,924	318,938 +60,257	14,530,000 +3,660,000	333 -54	5.1 -0.53

출처 오늘의 베스트 사이트(7월 기준)

순위	커뮤니티
1	dcinside.com
2	FM
3	Q
4	NATE 판
5	RULIWEB
6	1
7	PPOMPPU
8	INVEN
9	CLIEB.NET
10	인스타그램

#친환경 #환경오염

소비자들의 반응

감성사전 구축 개요



↓
1문장으로 나누어 감성 라벨링 – 감성 문장 추출
Mecab명사 중 불용어 라벨링 – 불용어로 문장 토큰화
TF-IDF로 전처리



2 감성사전 구축

데이터 전처리

(1) Clean_text 함수를 통해 영어, 기호를 정제하여 각각의 csv 파일 추출

예시 친환경_news.csv 중 한 문장

지난 6월에는 WPO(World Packaging Organization)가 개최한 ‘World Star Ceremony 2021’에서 재생 플라스틱과 재생지, 식물성 칫솔모, 국제삼림관리협의회(FSC) 인증 재료를 사용한 ‘그린클린’이 ‘Sustainable Packaging(지속가능한 포장)’ 부문에서 은상을 수상하기도 했다.



Clean_text 함수 사용

지난 6월에는 가 개최한 ‘ 2021’에서 재생 플라스틱과 재생지 식물성 칫솔모 국제삼림관리협의회 인증 재료를 사용한 ‘그린클린’이 지속가능한 포장’ 부문에서 은상을 수상하기도 했다.

“ 총 6개의 csv 파일 생성 ”

환경오염_news.csv
친환경_news.csv

환경오염_community.csv
친환경_community.csv

환경오염_sns.csv
친환경_sns.csv



2 감성사전 구축

데이터 전처리

(2) 수집된 3개의 csv를 1문장씩 '다'를 기준으로 나눈 후
 자른 하나의 문장을 -1(부정), 0(무의미), 1(긍정)으로 분류 - 3명이상판단

하면서 정작 여기저기 과대표장 남발로 환경오염 더 시키는건 생각안하누쓰래기백	-1	-1	0
현대백화점그룹은 경영 강화를 위해 전사적으로 친환경 경영활동을 펼치고 있다	1	1	1
고 말했다	0	0	0



-1,1로 라벨링된 문장으로 csv추출

labeling_data.csv

기후위기 대응법 에 번 등장한 녹색성장 제명에 녹색성장 이 들어간 이 법안에는 부칙을 포함해 총 번 녹색성장 (1
법안은 녹색성장을 경제와 환경이 조화를 이루는 성장 이라고 정의한다	1
제명에 녹색 이 붙긴 했지만 경제성장을 위한 수단으로서의 친환경기술 을 강조한다	1
년 월 당시 한국환경기술개발원이 낸 휴 폐광된 금속광산 지역의 오염 관리대책 을 보면 용화광산 주변 토양의 중	-1
광산 지역에서 떨어진 곳에서도 비소가 배 높았다	-1
안내판에 영풍 이란 이름에 주목한 건 이 기업의 석포제련소 환경오염 문제 때문이다	-1

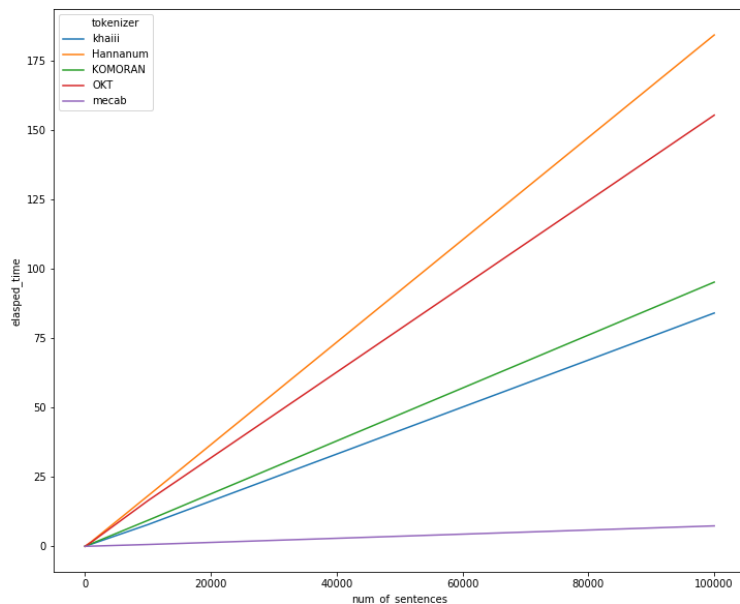
2 감성사전 구축

데이터 전처리

(3) 앞에서 생성한 6개의 csv파일을 Mecab을 이용하여 명사추출 후 불용어 라벨링

Mecab 형태소 분석기

사용 이유



분석 성능과 수행속도를 모두 갖춘
Mecab 형태소 분석기 사용

개발	642
기자	966
플라스틱	482
탄소	621
무단	217
청정	68
배포	186
효율	405
수거	217
금지	246
자연	187
태양광	248
전기	549
안전	360
빨대	77



개발	1
기자	0
플라스틱	1
탄소	1
무단	0
청정	1
배포	0
효율	1
수거	1
금지	1
자연	1
태양광	1

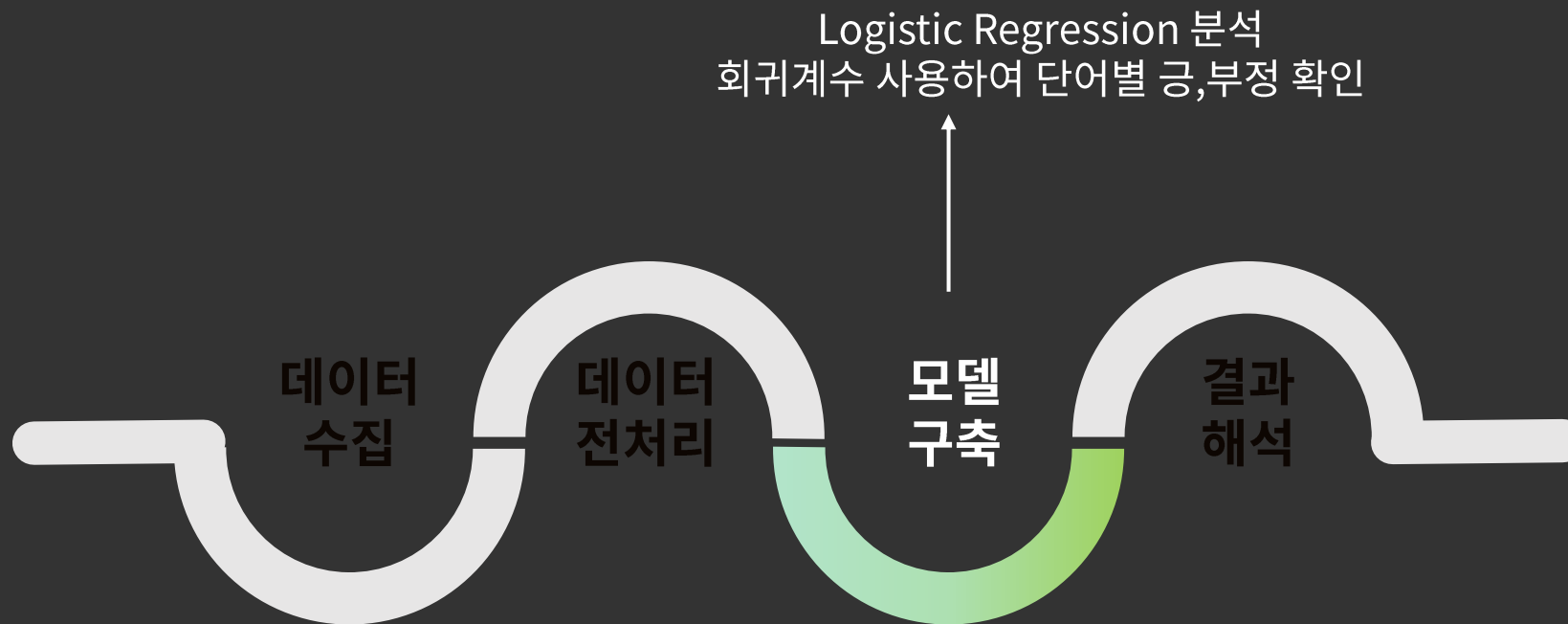
mecab.noun으로 명사 추출 후
불용어 라벨링 [0,1]

데이터 전처리

감성 [-1,1]인 문장만 추출한 `labeling_data.csv`를 앞에서 생성한 불용어를 제거함으로써 토큰화 이후 `tdm`에 `tf-idf` 가중치를 사용하여 데이터 행렬을 구성

[illegible]

감성사전 구축 개요



2 감성사전 구축

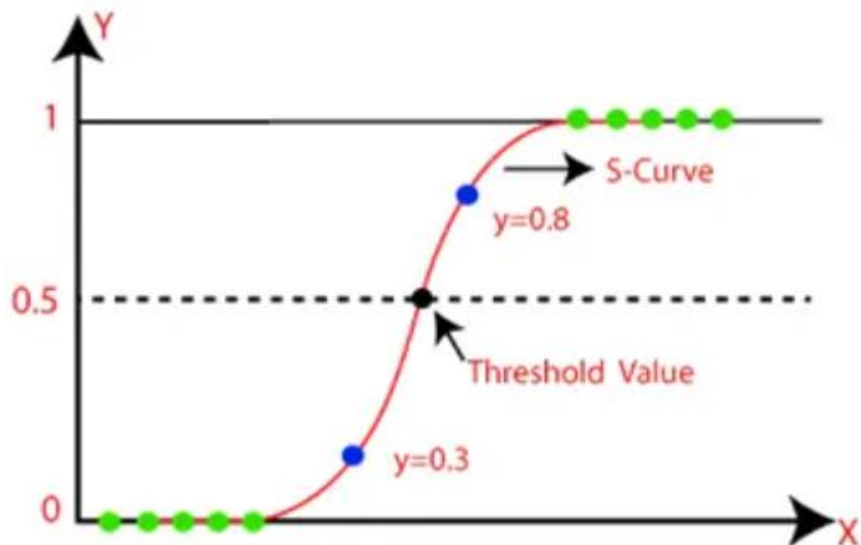
모델 구축

Logistic Regression의 회귀계수로
긍정 · 부정 단어 도출하여 환경 감성사전 구축

회귀계수로 단어별 긍부정 확인

Y (부정 -1, 긍정 1)

$$= (\text{단어}) \times \text{회귀계수} + (\text{단어}) \times \text{회귀계수} + \dots + b$$



예시문장

재활용을 하고 장바구니를 쓰고
기부도 하고 공정무역 제품을 산다

$$1 = \text{재활용} \times (1.696012) + \text{장바구니} \times (1.415834) \\ + \text{기부} \times (0.302357) + \text{공정무역} \times (-0.3768) \\ + \text{제품} \times (1.134818) + b$$

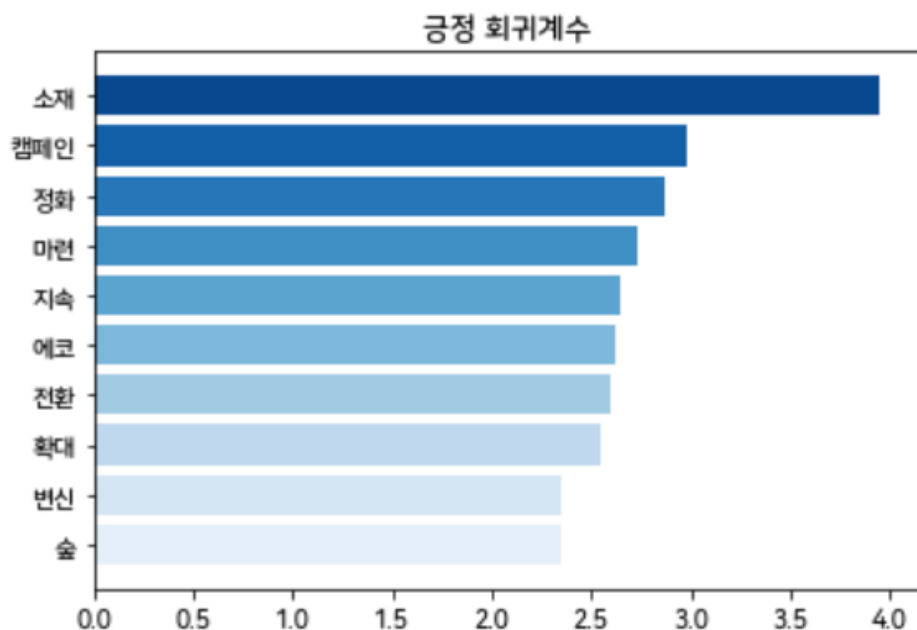
재활용	1.696012
장바구니	1.415834
기부	0.302357
공정무역	-0.3768
제품	1.134818

2 감성사전 구축

감성사전 구축 결과

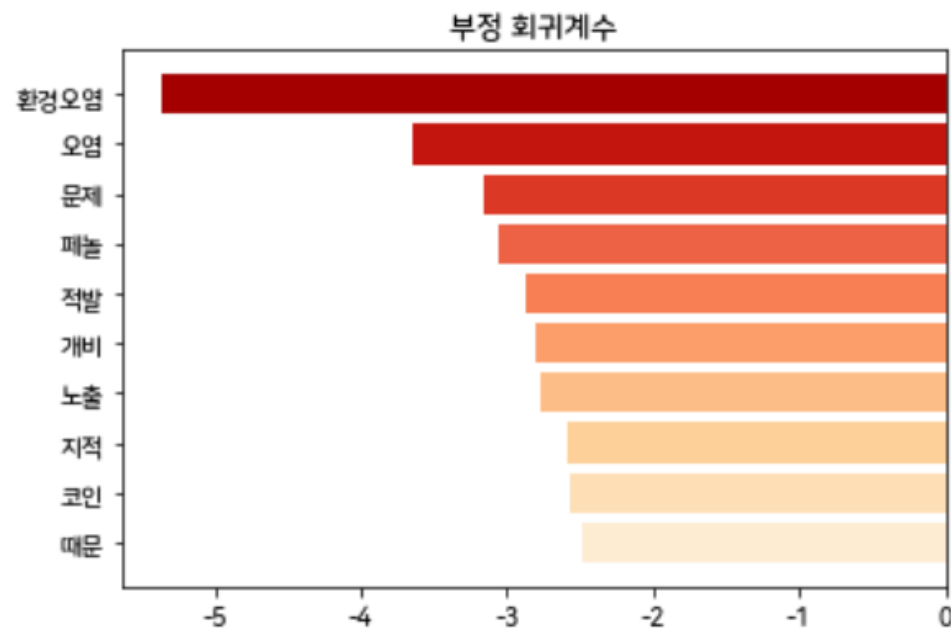
Logistic Regression의 회귀계수로
긍정 · 부정 단어 도출하여 환경 감성사전 구축

긍정 단어 top10



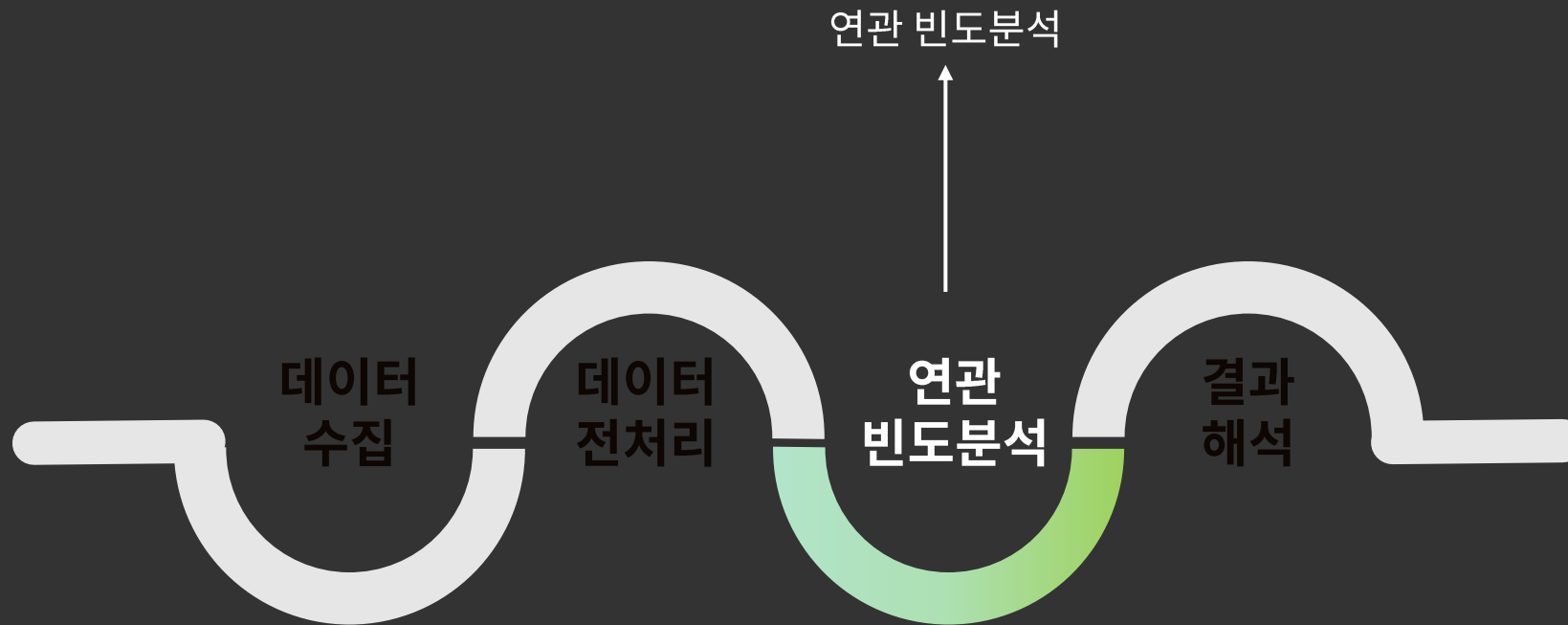
숲 에코 지속 정화 캠페인

부정 단어 top10



코인 적발 폐놀 오염

연관빈도분석 개요





연관빈도분석

연관빈도분석

6개의 csv파일을 전처리한 후 나온 명사들로
각 매체(communit, sns, news 별) 연관빈도분석을 진행

	word1	word2	freq
0	사업	환경	2068
1	친환경	환경	1836
2	오염	환경	1717
3	사업	친환경	1650
4	환경	환경오염	1645
5	배출	환경	1610
6	사용	친환경	1540
7	계획	사업	1497
8	사업	시설	1437
9	계획	환경	1430
10	배출	사업	1409
11	대한	환경	1357
12	지역	환경	1353
13	시설	환경	1345
14	사업	지역	1314
15	계획	친환경	1311
16	물질	환경	1286
17	생산	친환경	1277
18	사용	환경	1248

news_words_freq.csv

	word1	word2	freq
0	환경	환경오염	504
1	친환경	환경	349
2	친환경	플라스틱	296
3	플라스틱	환경	286
4	지구	환경	280
5	제로	플라스틱	265
6	플라스틱	환경오염	249
7	쓰레기	환경오염	227
8	제품	친환경	223
9	쓰레기	환경	223
10	지구	친환경	221
11	지구	환경오염	219
12	이스트	플라스틱	210
13	사용	환경	202
14	환경	활동	199
15	쓰레기	플라스틱	195
16	비건	환경오염	190
17	제로	친환경	187
18	바다	환경오염	185

sns_words_freq.csv

	word1	word2	freq
0	오염	환경	1228
1	친환경	환경	1133
2	사용	친환경	1126
3	사람	환경	1069
4	생각	친환경	1050
5	생각	환경	1024
6	사용	환경	1022
7	사람	생각	947
8	경제	한국	932
9	전기	친환경	919
10	에너지	태양광	912
11	사용	제품	910
12	사업	친환경	891
13	중국	환경	888
14	문제	환경	884
15	세계	환경	874
16	제품	친환경	872
17	정부	한국	866
18	세계	친환경	862

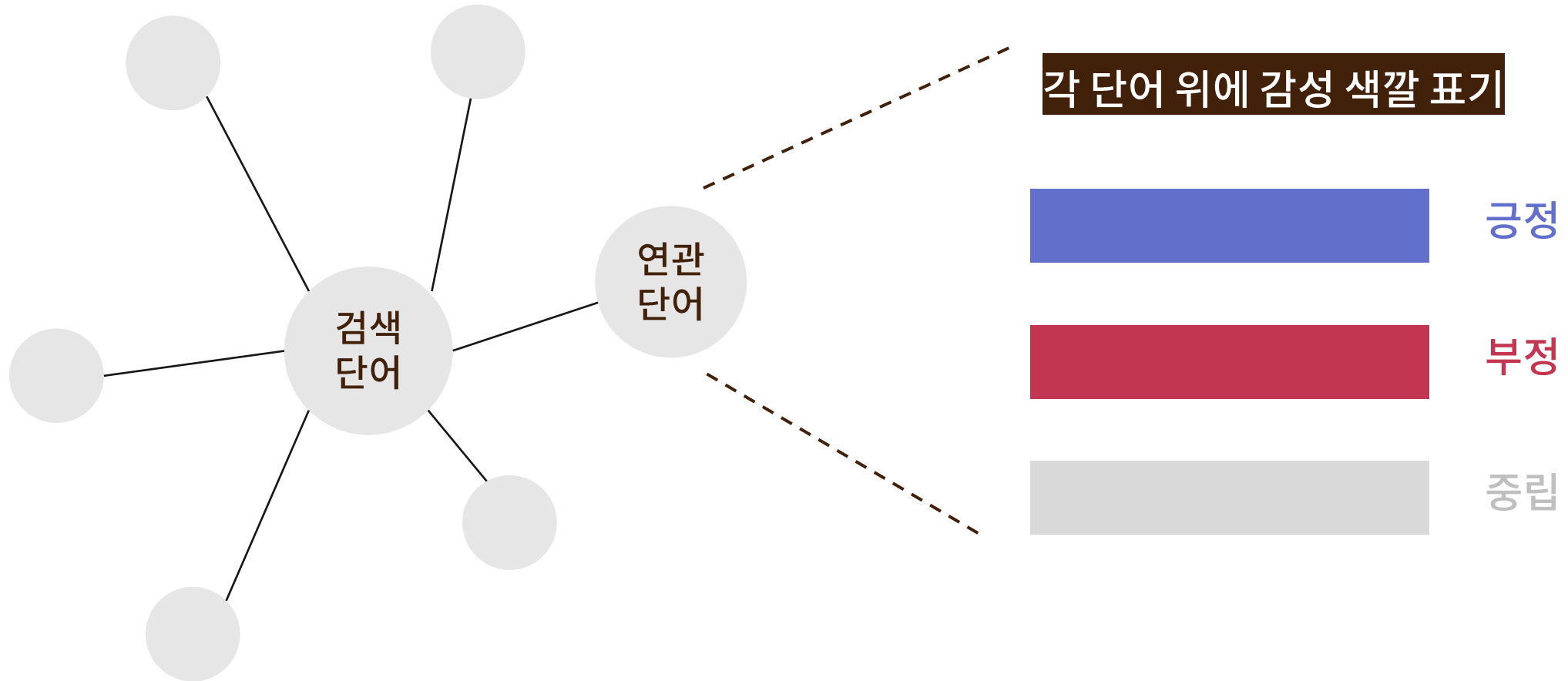
community_words_freq.csv



4 결과 해석

단어 Network 생성

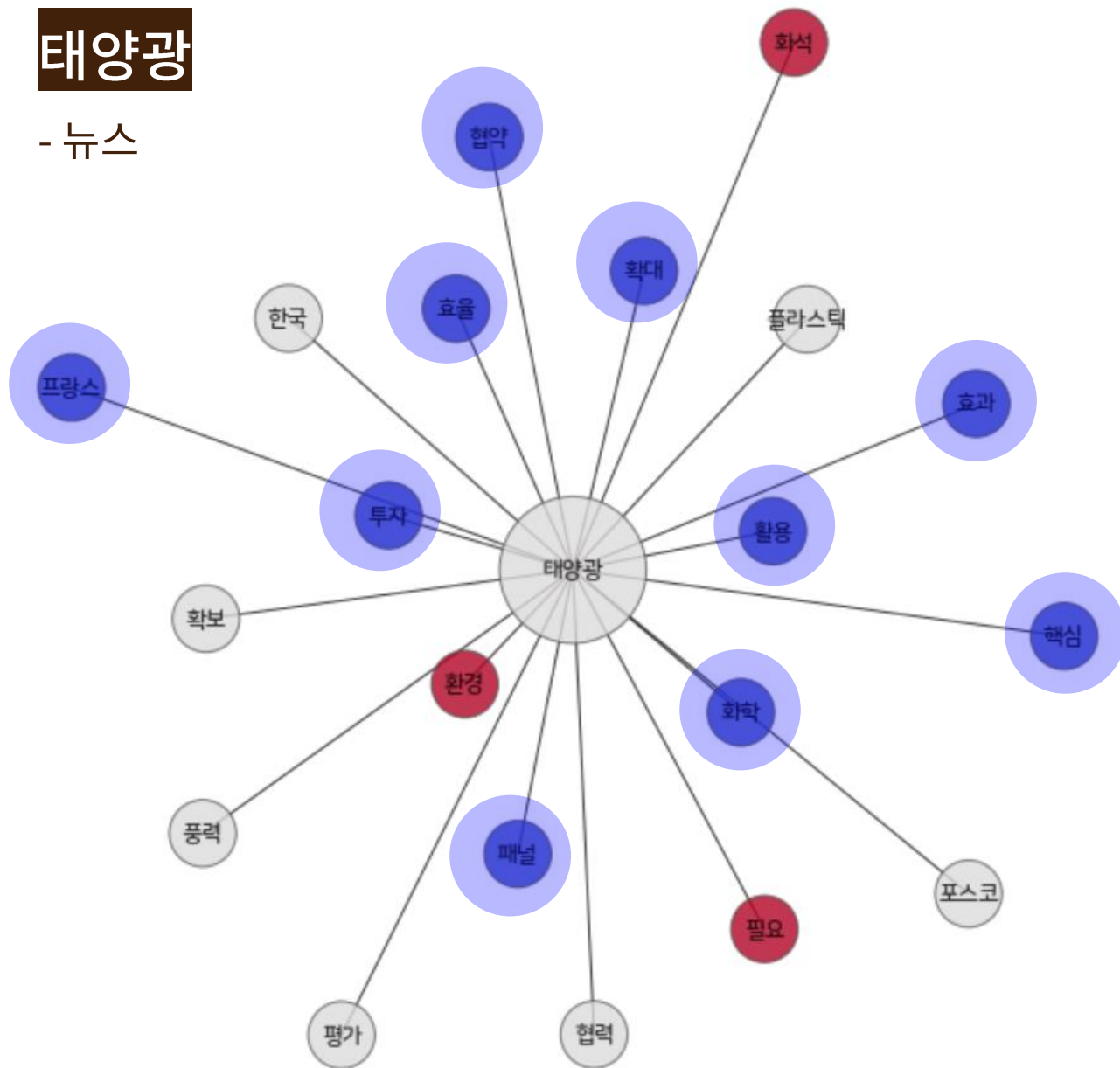
연관빈도분석으로 단어 network를 생성하며
해당하는 단어 위에 감정사전(dict.csv)을 활용하여 감정을 표기



4 결과 해석

태양광

- 뉴스



부정단어



긍정 단어

#효율

효과

활용

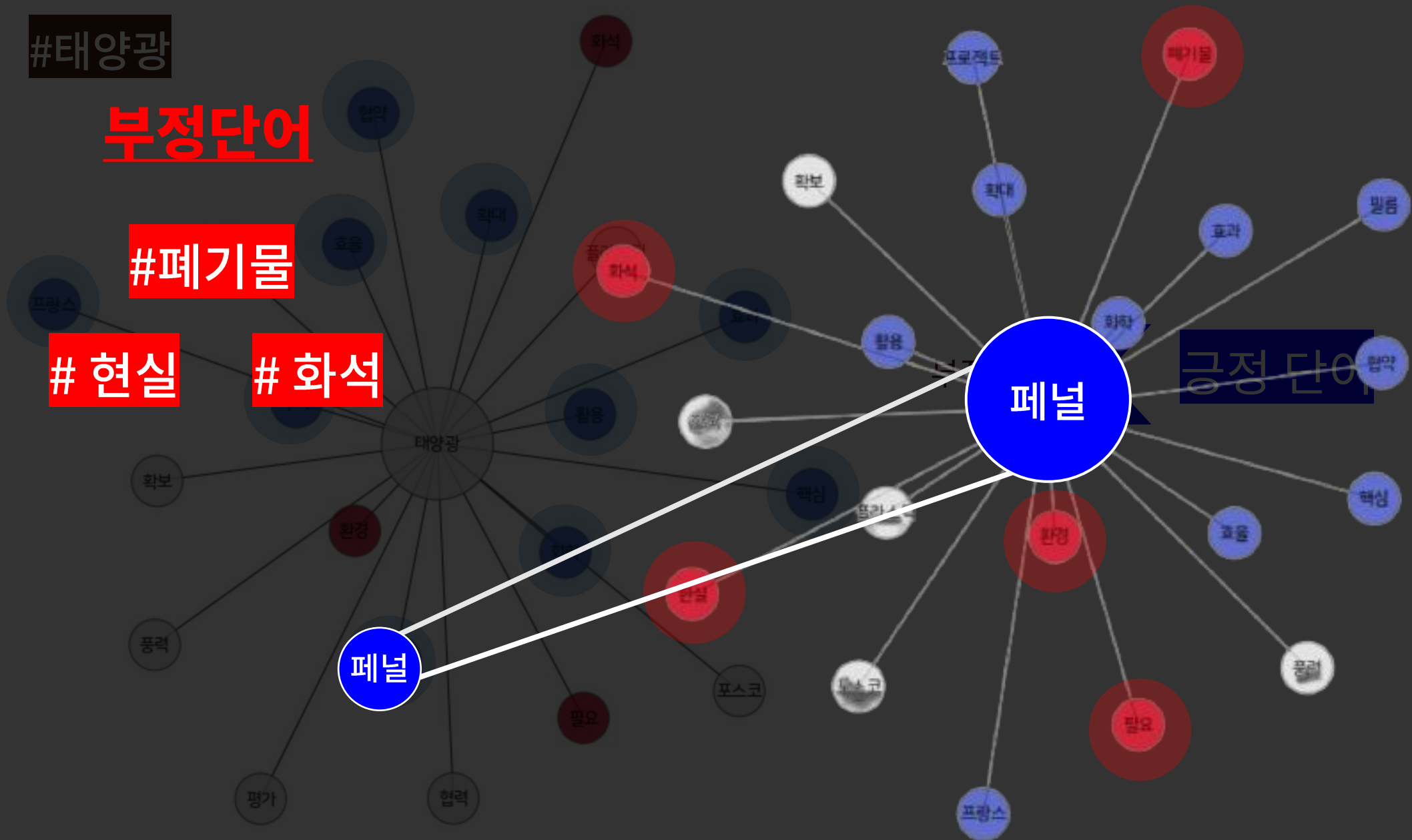
#태양광

#폐기물

현실

화석

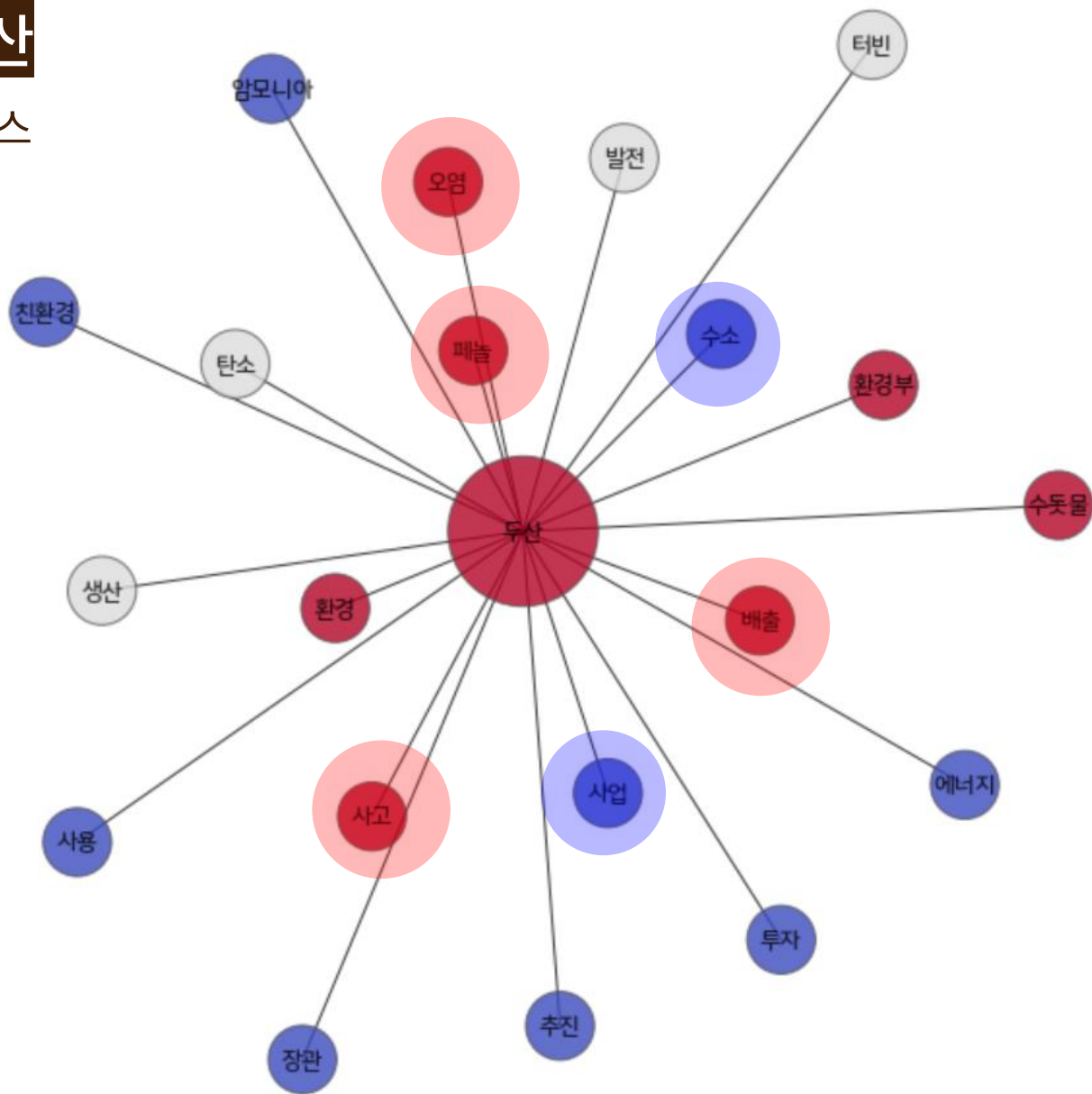
긍정 단어



4 결과 해석

두산

- 뉴스



부정 단어

페놀 # 사고

배출 # 오염

→ 낙동강 페놀 오염사고

긍정 단어

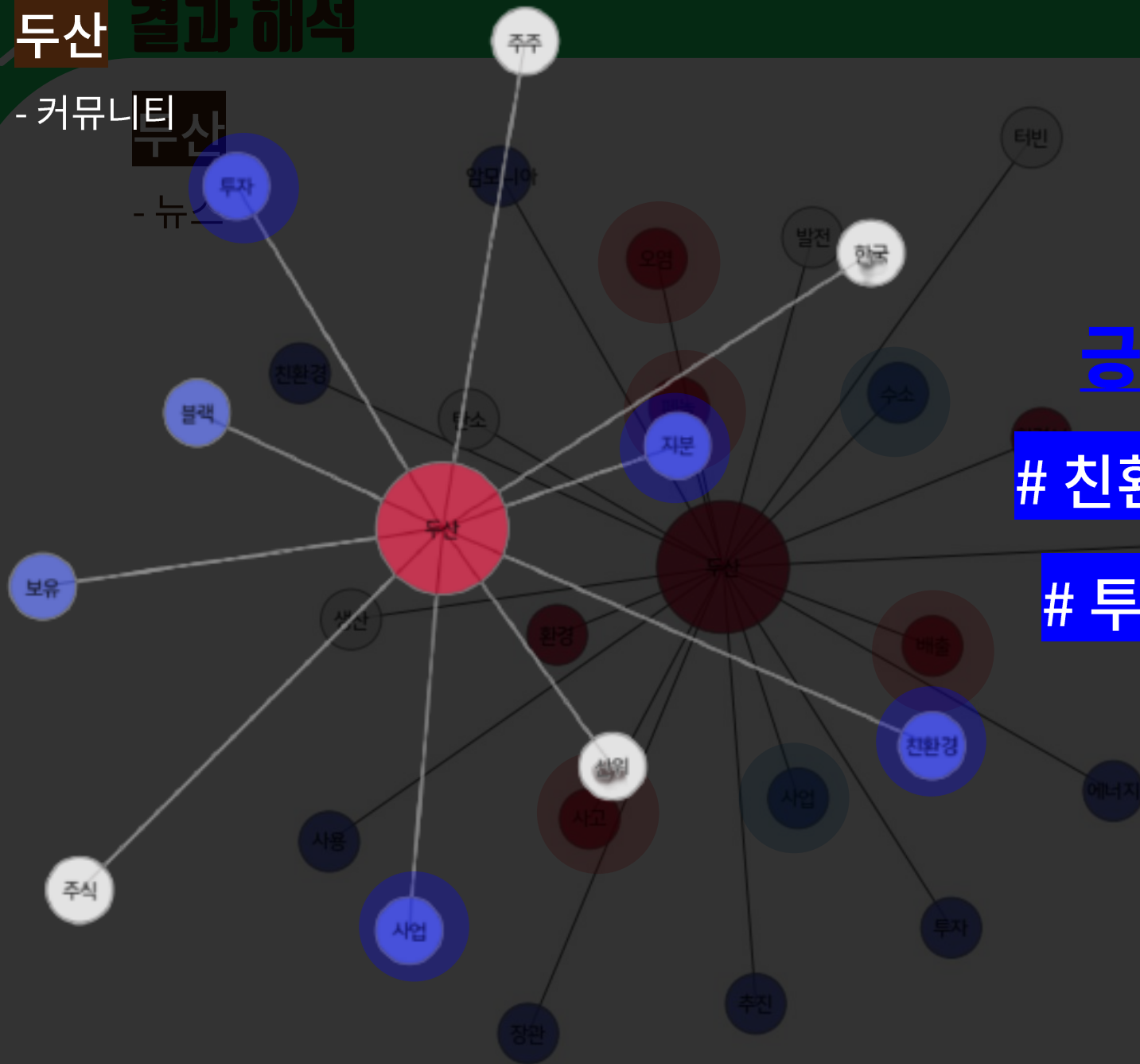
수소 # 사업

→ 두산의 수소시장 투자사업

- 커뮤니티

- 커뮤니티

두산

$$-\frac{L}{\pi}$$


친환경 # 지분

투자 # 사업

word1	word2
두산	지분
두산	블랙
두산	산위
두산	보유
두산	투자
두산	사업
두산	한국
두산	주주
두산	주식
두산	친환경

→ 두산의

4 결과 해석

테슬라 - 비트코인 - 전기차

비트코인

부정 단어

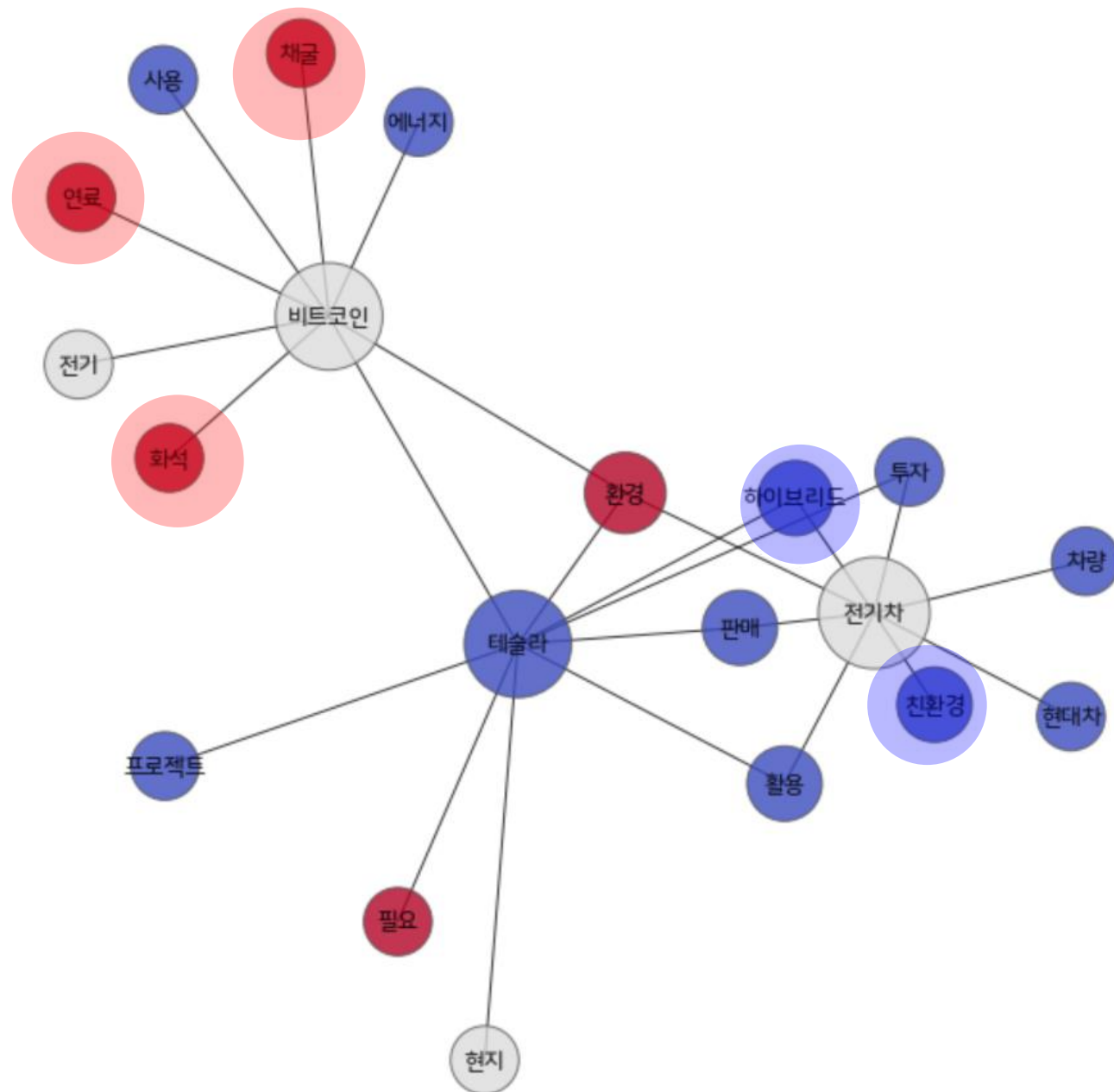
#채굴 #연료 #화석

→ '일론 머스크'의 잦은 언급으로 인한 결과

전기차

긍정 단어

친환경 # 하이브리드





5 활용방안 및 기대효과

활용 방안

· 친환경 마케팅에 사전 활용

환경감성사전을 참고하여
기업의 마케팅 활동 및 기획에 있어
긍부정 단어 활용 가능

· 진행 사업의 현황 파악

진행하고 있는 기업의
친환경 사업을 검색한 후
도출된 긍, 부정 단어 탐색으로
사업 현황 파악

기대 효과

기존 범용감성사전이 아닌,
환경 감성사전을 새로 구축함으로써
환경 관련 단어의 **높은 예측정확도**를
보여주는 환경 감성사전

새로 구축한 맞춤형 감성사전 사용시,
기존의 범용감성사전에 비해 평균
2.82%의 예측정확도의 향상 확인 가능

출처 [논문] 영역별 맞춤형 감성사전 구축을 통한
영화리뷰 감성분석



한계점

· 키워드 판별의 한계

1) 영어가 제외되어
LG전자 등이 검색되지 않음

2) '삼성전자' 인식하지만
'삼성 전자'의 경우
삼성 / 전자 를 별개로 인식

· 편향된 데이터

데이터 수집시
기간 설정을 하지 못해
데이터가 특정 시점에
편향될 가능성이 있음

· 감성사전의 객관성 부족

1) 수집된 데이터의 양이
많지 않아 결과가 제한적이며
일반화하기 어려움

2) 라벨링을 직접 하여
다소 주관적일 수 있음

감사합니다