

NO 1.

BERTopic\_2는 해당 링크 참고하여 실행. (<https://mz-moonzoo.tistory.com/23>)

토픽 번호	문서 수	주요 키워드	주제
-1	105	combustion, internal, engine, carbon, and...	내연기관/연소 관련 문서이나 군집되지 않은 문서 (키워드 다양도 높음, 방향성 모호)
0	75	filter, particulate, honeycomb, filtration, ceramic	DPF 및 미세입자 필터링 (허니컴 구조, 세라믹 필터 등)
1	61	exhaust, gas, temperature, pressure, engine	배기 가스 및 온도/압력 제어 시스템 (센서 기반 엔진 제어 가능성)
2	50	vehicles, equipment, automobiles, buildings	차량 및 장비 관련 일반 문서 (구체 주제 없음, 외연 넓음)
3	49	nox, reduction, emissions, burn, lean	질소산화물(NOx) 저감 (연소 최적화, lean burn 등 포함)
4	36	catalyst, catalytic, oxidation, deterioration	촉매 반응 및 열화 (산화 촉매의 열화 또는 수명 관련)
5	32	sulfur, removal, hydrogen, sulfide	황 제거 및 황화수소 처리 기술 (탈황 공정)
6	27	effects, harmful, environment, pollution	대기오염 및 환경 영향 분석 (건강 영향, 유해성 중심)
7	25	fuel, engines, engine, diesel, lean	연료 소비 및 디젤 엔진 운용 (연비, lean mixture 등)
8	19	nitrogen, oxides, removing, oxide	질소산화물 제거 기술 (SCR 등) (촉매제 중심)
9	12	ammonia, infrastructure, distribution, possible	암모니아 기반 인프라 시스템 (분산 및 저장 기술 포함)

코드: BERTopic

실행 결과(엑셀 sheet 이름): entity\_topics\_2

NO 2.

BERTopic\_3 version은 entity\_text에서 추출한 핵심 키워드를 중심으로 비슷한 것끼리 묶어 보는 걸 목표로 실행.

→ KeyBERT 기반으로 entity\_text를 핵심 키워드 중심으로 재구성하고, BERTopic으로 클러스터링

토픽 번호	문서 수	주요 키워드	주제
-1	112	increase, carbon, diesel, denitration, soot, ...	키워드가 흩어져 있거나 의미적으로 고립된 경우 (노이즈)
0	145	engine, combustion, exhaust, catalyst, diesel	엔진 연소 및 배기가스 관련 기술 (내연기관, 촉매 반응 등)
1	65	nox, nitrogen, oxides, reduction, catalyst	질소산화물(NOx) 저감 관련 기술 (촉매 환원 등)
2	63	filter, honeycomb, filtration, particulate	DPF/PM 필터링 구조 (허니컴 필터, 세라믹 입자 제거 등)
3	32	sulfur, removal, hydrogen sulfide	황 제거 기술 (H <sub>2</sub> S 제거, 황화물 처리 등)
4	27	harmful, environment, pollution, air	환경 유해성 및 대기오염 영향 분석 (오염원, 건강 영향)
5	20	vehicles, automobiles, military, equipment	차량/군용 장비 관련 문서 (배기 관련보다 차량 자체에 초점)
6	14	ammonia, infrastructure, release, precursor	암모니아 기반 배출 시스템 및 인프라 구축 논의
7	13	manufacturers, researchers, commercially	제조사/연구자 중심의 연구 동향 또는 기술 상용화 문서

코드: BERTopic\_keybert

파라미터: top\_n=3

실행 결과(엑셀 sheet 이름): entity\_topics\_keybert\_3

NO 3.

BERTopic\_4 version은 기존 'top\_n=3'에서 'top\_n=5'로 증가시켜 진행.

→ 비슷한 개념끼리 더 세분화되게 묶어보는 것을 목표로 실행.

즉, KeyBERT 기반으로 문장에서 top 5 핵심 키워드를 추출해 문서를 재구성.

이를 BERTopic에 입력하여 의미 표현을 확장한 상태에서 클러스터링 수행.

토픽 번호	문서 수	대표 키워드	주제
-1	96	soot, carbon, denitration, regeneration	탈질·탄소 관련 문서 중 노이즈 처리된 문서들
0	69	filter, particulate, honeycomb, ceramic	DPF/미세먼지 필터링 및 허니컴 세라믹 구조
1	60	exhaust, gas, temperature, flow	배기 가스 온도·유량 제어 및 정화 장치
2	47	nox, reduction, adsorber, emissions	질소산화물(NOx) 흡착·저감 시스템
3	39	catalyst, oxidation, deterioration	산화 촉매 및 열화 현상
4	33	engine, fuel, diesel, lean	연료소모, 디젤엔진, 희박연소(lean burn)
5	30	effects, harmful, environmental	환경유해성 및 공해 영향 분석
6	30	sulfur, hydrogen, sulfide, removal	황 제거 및 황화수소 처리 단계
7	20	nitrogen, oxides, gases, reduction	질소산화물 제거/환원 (SCR 포함 가능)
8	19	vehicles, automobiles, owner, military	차량 및 군수 장비 관련 문서
9	19	internal, combustion, engine	내연기관 기술 및 연소 원리
10	16	manufacturers, researchers, commercial	산업계 및 연구기관 관련 기술 동향
11	13	ammonia, infrastructure, distribution	암모니아 기반 연료 인프라 및 배포 문제

코드: BERTopic\_keybert

파라미터 변경: top\_n=5

실행 결과(엑셀 sheet 이름): entity\_topics\_keybert\_4