

연관 네트워크 분석

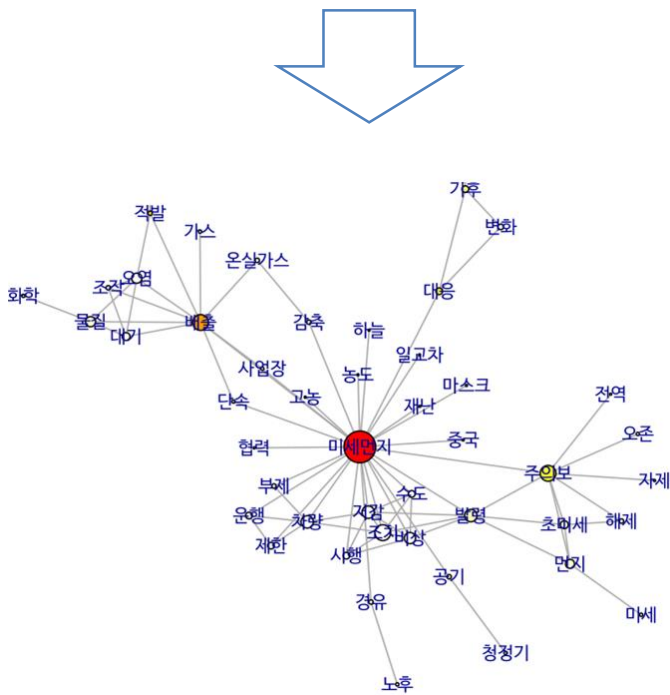


한국환경정책·평가연구원
환경경제연구실
부연구위원 진대용

키워드 네트워크 분석 서비스 개요

- 키워드 네트워크 분석 서비스
 - 사용자가 입력한 파일에 대해 네트워크 구성 One-Stop 서비스
 - 환경 텍스트의 키워드를 파악하고 키워드 사이의 연관 빈도가 높은 키워드들의 네트워크를 도출

DATE	TITLE	CATEGORY
19960206	우리나라 환경관련 법안정책의 개선방안	정책일반
19960109	산업별 공업용수의 수요·수급·수질원형파악 및 재이용에 따른 제반문제 검토에 관한 연구	물환경
19960109	중립환경정보망 개발사업(III)	정책일반
19970502	"지방의제21" 모델 개발연구	정책일반
19950127	오염지표성질을 이용한 연안환경관리	물환경
19950127	직접여과법	물환경
19950127	다이옥신에 관한 검토 및 분석기술	대기환경
19950127	환경개선 부담금제도 개선방안	환경정책
19950127	환경기술 연구개발 관리체계 구축방안(I)	정책일반
19950127	환경기술 연구개발 관리체계 구축방안(II)	정책일반
19950127	환경기술 연구개발 관리체계 구축방안(III)	정책일반
19950127	한국형 선진환경산업의 육성정책 개발을 위한 기초조사(I)	정책일반
19950127	한국형 선진환경산업의 육성정책 개발을 위한 기초조사(II)	정책일반
19950127	수질중립규제방식의 활용방안에 관한 연구(I)	물환경
19950127	국제환경협약의 국내 대응기반 연구(I)	정책일반
19950127	최근전해성 평가방법의 개선과 향후연구에 관한 연구	정책일반



연관 분석

- SUPPORT(지지도)
 - 키워드 A, B가 같은 문서에서 동시에 발생할 $P(A \cap B)$
- CONFIDENCE(신뢰도)
 - 키워드 A가 나타났을 때, 키워드 B가 같은 문서 내에서 발생 $P(B|A)$
- LIFT(향상도)
 - 키워드 A와 B가 동시에 나타날 확률(지지도)값에 각각의 키워드가 나타날 문서에 확률을 나누어 주는 패널티로서 $P(A \cap B)/P(A)P(B)$

〈표 2-3-14〉 NAVER 환경뉴스(2013~2016년) 키워드 연관 분석

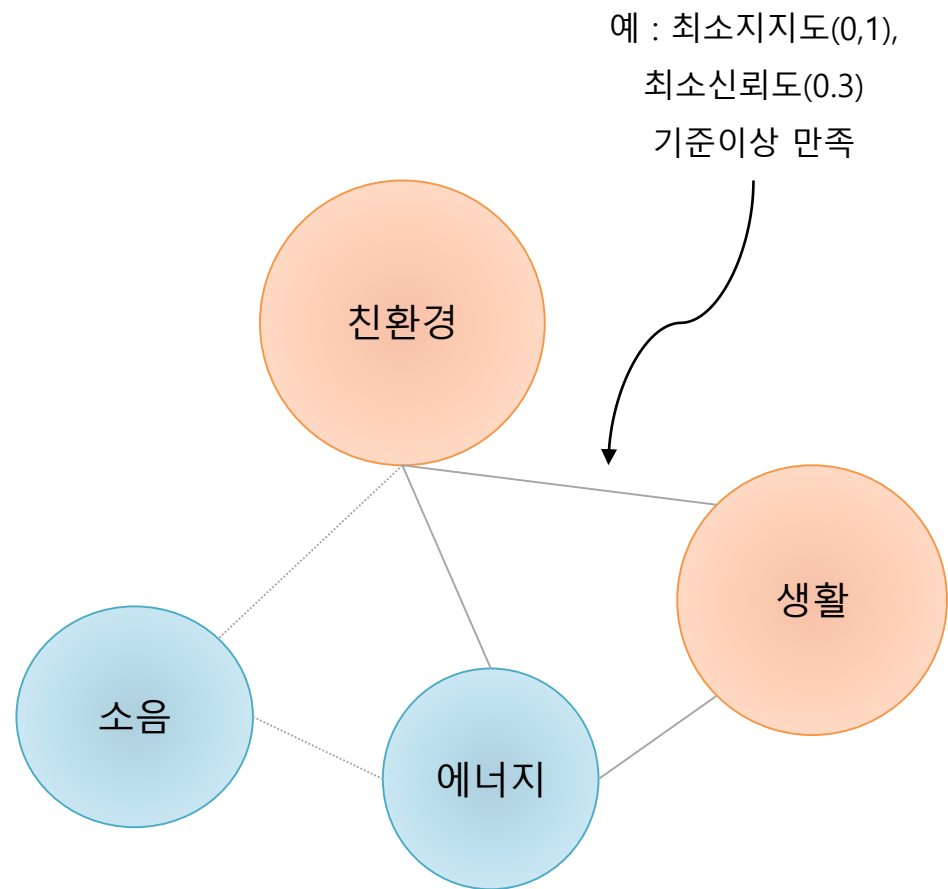
순번	lhs		rhs	지지도	신뢰도	향상도
1	간접	⇒	태풍	0.002	0.960	51.279
2	태풍	⇒	간접	0.002	0.081	51.279
3	대설	⇒	특보	0.001	0.380	40.323
4	특보	⇒	대설	0.001	0.128	40.323
5	최강	⇒	한파	0.001	0.688	62.212
6	한파	⇒	최강	0.001	0.105	62.212
7	고니	⇒	태풍	0.001	0.675	36.055
8	태풍	⇒	고니	0.001	0.061	36.055
9	반달	⇒	지리	0.001	0.685	153.626
10	지리	⇒	반달	0.001	0.237	153.626
11	가슴	⇒	지리	0.001	0.667	149.529
12	지리	⇒	가슴	0.001	0.237	149.529

자료: 저자 작성.



키워드 네트워크 시각화

- 키워드들 사이의 연관성을 계산하여 네트워크 시각화
- 예 : '친환경' 네트워크 (그래프)
 - 노드와 엣지는 연구자의 판단에 따라 여러 방법으로 정의할 수 있음
 - 노드 (node)
 - 노드 크기 : In-Degree (노드로 들어오는 엣지의 개수)
 - 노드의 색깔 : Betweenness Centrality
 - 엣지 (edge)
 - 지지도와 신뢰도가 특정 기준 이상일 경우 엣지



키워드 네트워크 분석 과정

입력 파일 포맷 (XLSX)

DATE	TITLE	CATEGORY
19960206	우리나라 환경관련 예산정책의 개선방안	정책일반
19960109	산업별 공업용수의 수요·수질·수질현황파악 및 재이용에 따른 제반문제 검토에 관한 연구	물환경
19960109	중합환경정보망 개발사업(III)	정책일반
19970502	「지방의제21」 모델 개발연구	정책일반
19950127	오염지표생물을 이용한 연안환경관리	물환경
19950127	직접여과법	물환경
19950127	다이옥신에 관한 검토 및 분석기술	대기환경
19950127	환경개선 부담금제도 개선방안.	환경정책
19950127	환경기술 연구개발 관리체계 구축방안(I)	정책일반
19950127	환경기술 연구개발 관리체계 구축방안(II)	정책일반
19950127	환경기술 연구개발 관리체계 구축방안(III)	정책일반
19950127	한국형 선진환경산업의 육성책 개발을 위한 기초조사(I)	정책일반
19950127	한국형 선진환경산업의 육성책 개발을 위한 기초조사(II)	정책일반
19950127	수질총량규제방식의 활용방안에 관한 연구(I)	물환경
19950127	국제환경협약의 국내 대응기반 구축	정책일반
19950127	환경적합성 평가기법의 개선과 활용방안 연구	정책일반
19950127	환경적합성 평가방법의 개선 및 활용방안연구	정책일반
19950127	지방자치단체의 환경예산분석과 환경적합성 평가에 관한 연구	환경정책
19950127	환경투자재원 조달에 관한 연구	정책일반
19950127	중합환경정보망 개발사업	정책일반
19950127	지방자치시대의 환경정책	정책일반
19950127	OECD 가입과 한국의 환경정책 개선방향	정책일반
19950127	용융수관리법(안)의 문제점 및 개선방향	물환경
19950127	음식물쓰레기 감량화 규제에 대한 연구	폐기물관리

키워드 네트워크 분석 웹서비스

키워드 연관 네트워크 분석

역설 파일 선택(*.xlsx)

Browse... kin_test3.xlsx

Upload complete

기타고리

정책일반

시작날짜

20100101

끝날짜

20171231

필터링 실행

시도

1001

지치도

0.01

신뢰도

0.3

관계수

30

문헌어

2개 이상 관계 있음

형태소 분석기

한나눔

의역소기

정렬방법

URL

Support

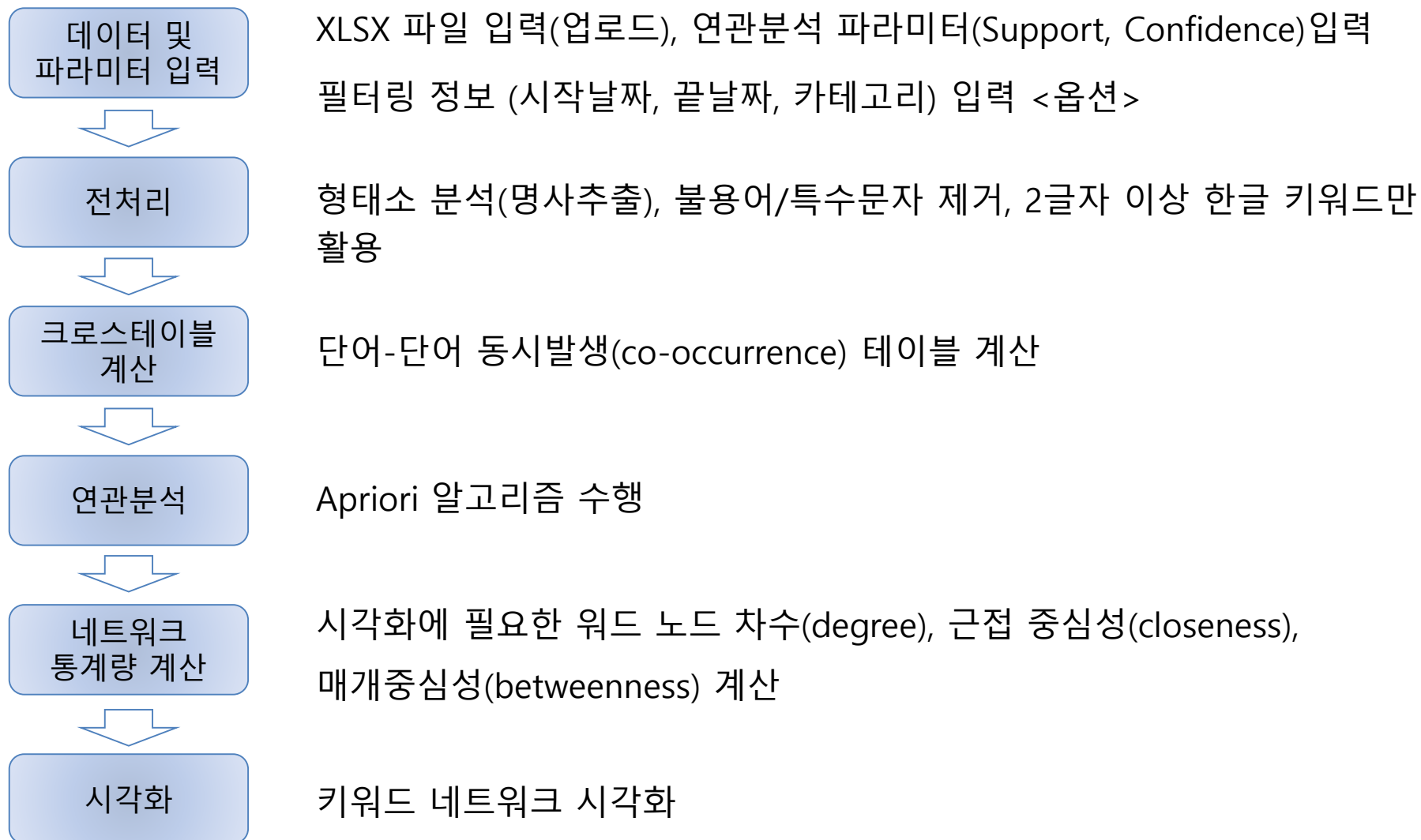
데이터 연관분석 네트워크 시각화 네트워크 지표 매뉴얼

Show 50 entries

Search:

	date	text	category
1	19960206	우리나라 환경관련 예산정책의 개선방안	정책일반
2	19960109	산업별 공업용수의 수요·수질·수질현황파악 및 재이용에 따른 제반문제 검토에 관한 연구	물환경
3	19960109	중합환경정보망 개발사업(III)	정책일반
4	19970502	「지방의제21」 모델 개발연구	정책일반
5	19950127	오염지표생물을 이용한 연안환경관리	물환경
6	19950127	직접여과법	물환경
7	19950127	다이옥신에 관한 검토 및 분석기술	대기환경
8	19950127	환경개선 부담금제도 개선방안.	환경정책
9	19950127	환경기술 연구개발 관리체계 구축방안(I)	정책일반
10	19950127	환경기술 연구개발 관리체계 구축방안(II)	정책일반
11	19950127	환경기술 연구개발 관리체계 구축방안(III)	정책일반
12	19950127	한국형 선진환경산업의 육성책 개발을 위한 기초조사(I)	정책일반
13	19950127	한국형 선진환경산업의 육성책 개발을 위한 기초조사(II)	정책일반
14	19950127	수질총량규제방식의 활용방안에 관한 연구(I)	물환경
15	19950127	국제환경협약의 국내 대응기반 구축	정책일반
16	19950127	환경적합성 평가기법의 개선과 활용방안 연구	정책일반
17	19950127	환경적합성 평가방법의 개선 및 활용방안연구	정책일반
18	19950127	지방자치단체의 환경예산분석과 환경적합성 평가에 관한 연구	환경정책
19	19950127	환경투자재원 조달에 관한 연구	정책일반
20	19950127	중합환경정보망 개발사업	정책일반
21	19950127	지방자치시대의 환경정책	정책일반
22	19950127	OECD 가입과 한국의 환경정책 개선방향	정책일반
23	19950127	용융수관리법(안)의 문제점 및 개선방향	물환경
24	19950127	음식물쓰레기 감량화 규제에 대한 연구	폐기물관리

키워드 네트워크 구성 과정



키워드 네트워크 분석 서비스 활용 사례

- 사례1 : 네이버 환경뉴스 시각화

- **활용 데이터**

- 네이버 뉴스 제목 (2019년), 약 3만건

- 미세먼지

- 고농도 미세먼지, 중국, 공기청정기, 저감, 마스크, 차량 N부제, 노후경유차 등

- 기후변화

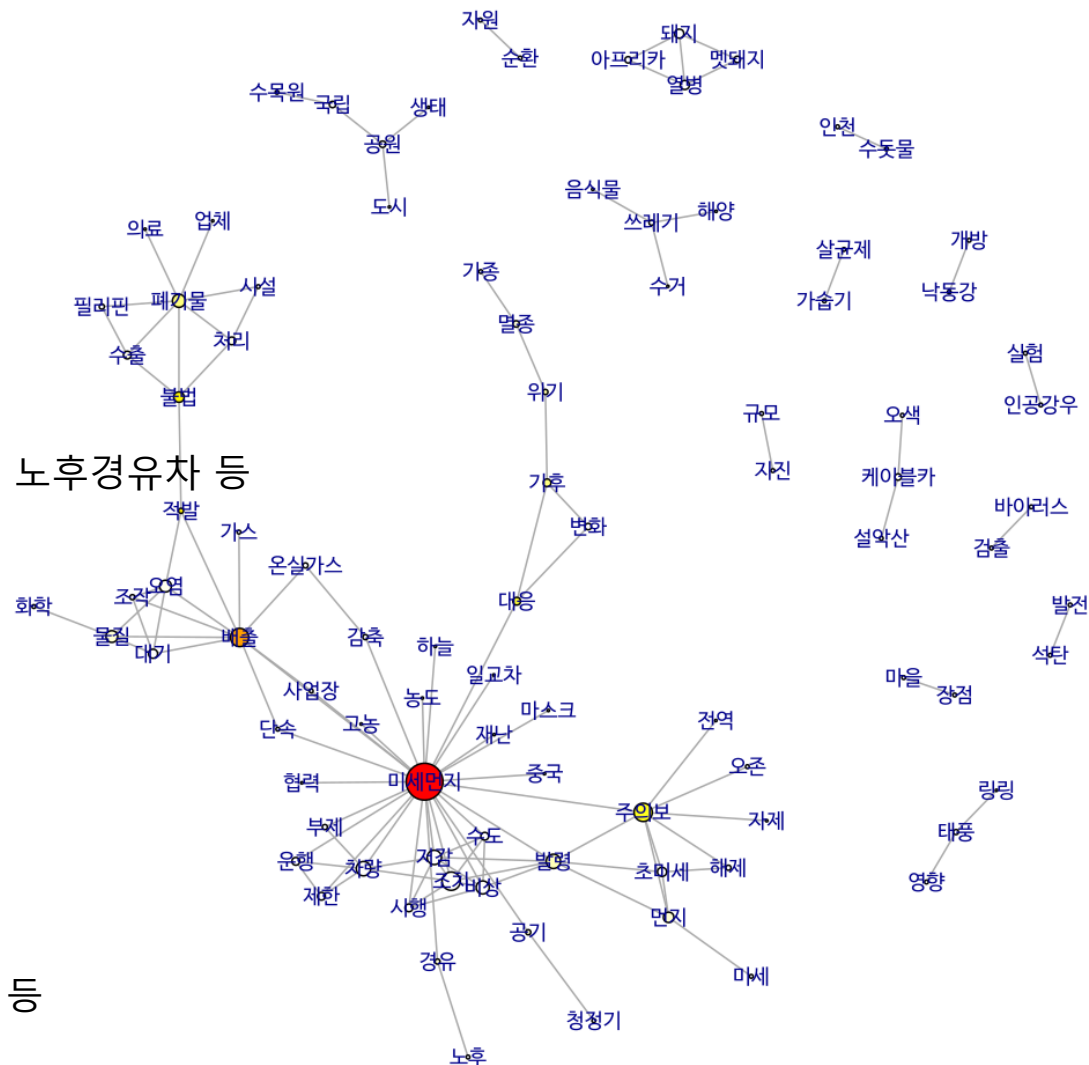
- 온실가스 배출, 감축, 대응, 멸종위기종 등

- **쓰레기/폐기물**

- 처리, 수출, 시설, 업체, 해양 쓰레기 등

- 기타

- 인천 수돗물, 석탄발전소, 태풍 링링, 돼지열병, 가습기 살균제 등



키워드 네트워크 분석 서비스 활용 사례

데이터	연관분석		네트워크 시각화	네트워크 지표	메뉴얼	
lhs	->	rhs	support	confidence	lift	count
{저감}	=>	{미세먼지}	0.024704200	0.8016878	5.147803	760
{미세먼지}	=>	{저감}	0.024704200	0.1586308	5.147803	760
{조미세}	=>	{먼지}	0.018755688	0.9914089	35.630496	577
{먼지}	=>	{조미세}	0.018755688	0.6740654	35.630496	577
{조치}	=>	{저감}	0.014562476	0.8028674	26.054232	448
{저감}	=>	{조치}	0.014562476	0.4725738	26.054232	448
{조치}	=>	{미세먼지}	0.014172409	0.7813620	5.017287	436
{변화}	=>	{기후}	0.011929528	0.9291139	41.008982	367
{기후}	=>	{변화}	0.011929528	0.5265423	41.008982	367
{비상}	=>	{조치}	0.011929528	0.6963947	38.394061	367
{조치}	=>	{비상}	0.011929528	0.6577061	38.394061	367
{비상}	=>	{미세먼지}	0.011409440	0.6660342	4.276743	351
{대기}	=>	{오염}	0.011181901	0.5880342	22.223936	344
{오염}	=>	{대기}	0.011181901	0.4226044	22.223936	344
{열병}	=>	{돼지}	0.010954362	1.0000000	76.910000	337
{돼지}	=>	{열병}	0.010954362	0.8425000	76.910000	337

데이터	연관분석	네트워크 시각화	네트워크 지표	메뉴얼			
연결중심성	값	근접중심성	값.1	매개중심성	값.2	고유중심성	값.3
미세먼지	0.28	미세먼지	0	미세먼지	0.294	미세먼지	1
배출	0.108	배출	0	배출	0.156	조치	0.553
주의보	0.086	사업장	0	적발	0.09	저감	0.553
저감	0.075	단속	0	불법	0.08	비상	0.499
조치	0.075	주의보	0	주의보	0.064	발령	0.472
폐기물	0.075	발령	0	대응	0.06	차량	0.418
발령	0.075	대응	0	폐기물	0.038	시행	0.387
비상	0.065	저감	0	기후	0.037	수도	0.387
차량	0.065	조치	0	위기	0.025	주의보	0.29
오염	0.054	비상	0	발령	0.02	배출	0.269
조미세	0.043	차량	0	먼지	0.013	제한	0.248
먼지	0.043	수도	0	물질	0.013	운행	0.248
대기	0.043	시행	0	멸종	0.013	부제	0.211
불법	0.043	공기	0	공기	0.013	사업장	0.189
물질	0.043	운행	0	경유	0.013	단속	0.189
수도	0.043	제한	0	처리	0.006	감축	0.158
시행	0.043	경유	0	수출	0.006	대응	0.157

키워드 네트워크 분석 서비스 활용 사례

- 사례 2 : KEI 연구보고서 네트워크

- **활용 데이터**

- KEI 연구보고서 제목 285건 (2018년~2019년)

- 기후

- 에너지, 온실가스
- 녹색성장, 경제

- 대기오염, 유해화학 물질

- 건강영향
- 배출시설저감
- ...

- 지속가능발전, 자원순환, 폐기물처리, 원자력 해체, 공공참여 등

- ...

