

TEXTOM is New Things

# 텍스톰SV 매뉴얼

10th Anniversary Special Version

## All in One, TEXTOM

다른 툴 필요없이 수집, 정제, 분석, 시각화 모두 가능합니다.

더욱 다양해진 분석툴로 더 면밀한 분석을 이용해보세요.

내 모든 현황을 한눈에 볼 수 있는 대시보드 제공, 더욱 편리해졌습니다.

# CONTENTS

목차를 클릭하면 해당 페이지로 넘어갑니다.  
이용 전 매뉴얼을 숙지하시면 쉽게 이용하실 수 있습니다.

intro. 로그인 / 입장

→ 바로 이동하기

1. 대시보드 (대시보드 개괄)

→ 바로 이동하기

2. 데이터수집

→ 바로 이동하기

3. 수집리스트

→ 바로 이동하기

4. 데이터 전처리

→ 바로 이동하기

5. 키워드 정제

→ 바로 이동하기

6. 데이터 분석

→ 바로 이동하기

7. 단어 분석

→ 바로 이동하기

8. 매트릭스

→ 바로 이동하기

9. 네트워크 분석

→ 바로 이동하기

10. 토픽 분석

→ 바로 이동하기

11. 감성 분석

→ 바로 이동하기

12. 가설 검증

→ 바로 이동하기

13. 커스터 마이징

→ 바로 이동하기

## 빅데이터를 쉽고 빠르게!

분석, 시각화, 인사이트를 다양하게 활용하세요.

1

TEXTOM >

TEXTOM Edu >

TEXTOM CHINA >

수집데이터

957,460건

분석데이터

106,206건

텍스톰회원수

22,784명

활용 논문수

309건

ver. 6.0 (2023.03월 현재)

## 텍스톰 가격에 대한 모든 정보 이용하기 전 꼭 살펴보세요!

데이터 구매 및 크롤링 등의 비용정보  
구매 시 필요한 양 산정 팁을 알려드립니다.

자세히 보기

## Bigdata Academy Online

단 4시간의 강의로 기초부터 심화까지!

바로가기 →

## AI의 원리를 배우는 AI EDUTOM

이해하고, 배우고, 만들어볼 수 있는 AI 교육플랫폼

① [TEXTOM] 버튼을 클릭하여 입장해주세요.

※ 로그인 및 회원가입은 각 버튼 클릭 후 페이지로 입장 시 가능합니다.

**1** 총 수집/분석현황  
총 수집 31159건 / 31158건

**2** 수집진행 현황  
30652 / 30653 포털/SNS 데이터  
0 / 0 뉴스 데이터  
506 / 506 보유 데이터  
1206 / 3853 정제/분석리스트

**3** 수집완료 리스트  
\* 완료된 최근 5개의 수집리스트만 보입니다. 전체 31158건

키워드	수집유형	용량
오마이오아이 oiioi +추천 -전시회 -...	포털/SNS	0 MB
오마이오아이 oiioi +추천 -전시회 -...	포털/SNS	0 MB
오마이오아이 oiioi +추천 -전시회 -...	포털/SNS	0 MB
오마이오아이 oiioi +추천 -전시회 -...	포털/SNS	0 MB
14 k-contents raw.txt	보유데이터	3.89 MB

**4** 데이터 전처리 리스트  
\* 진행중인 최근 5개의 수집리스트만 보입니다. 완료 0건 / 대기 1건

키워드	수집유형	수집상태
"마우스" -키보드	포털/SNS	대기 75 번

**5** theimc\_big 님 azurebeluga@theimc.co.kr

**6** 공지사항  
공지 <제 4기 TEXTOM Analyst>모집 안내 : 더욱 확대된 활동혜... 23.03.10  
공지 ★3월 31일 잠시 운영 중단 안내 23.03.03

**7** 시각화 리스트  
0320 테스트 23.03.21

- 1 업로드, 수집, 정제 등 텍스트용 이용현황을 알 수 있습니다.
- 2 수집 진행 중인 데이터를 확인할 수 있습니다.
- 3 수집이 완료된 데이터를 볼 수 있습니다..
- 4 특정 데이터셋의 전처리가 진행 중인지, 완료인지 확인할 수 있습니다.
- 5 잔여 데이터를 확인할 수 있습니다.
- 6 공지사항, 텍스트용 매뉴얼, 활용논문 등 정보를 확인할 수 있습니다.
- 7 생성한 시각화 데이터를 확인할 수 있습니다.

1

포털/채널

2

뉴스

3

보유데이터

키워드 미리보기

확인


수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

수집키워드

키워드

[키워드추가](#)를 사용하면 여러개의 [수집리스트](#)를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)

기간

 2023-02-27

~

 2023-03-06

1주

네이버 학술정보전체, 다음 웹문서, 트위터, 페이스북, 유튜브는 기간 설정이 불가능합니다.

수집단위

사용  사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 1 [포털/SNS] 채널 수집

- 네이버, 다음, 구글, 바이두, 유튜브, 트위터, 페이스북 데이터 수집이 가능합니다.

## 2 [뉴스] 채널 수집

- KBS, MBC, SBS, YTN, 조선일보, 중앙일보 등 언론사 20곳의  
기사 데이터 수집이 가능합니다. (댓글 데이터는 수집하지 않습니다)

## 3 [보유데이터] 수집

- 텍스트로 작성된 PDF, TXT, XLSX 형식의 파일 업로드가 가능합니다.

포털/채널

뉴스

보유데이터

1

키워드 미리보기

확인

수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

2

수집키워드

연관키워드

빅데이터   글로벌빅데이터   통계자료   메가스터디빅데이터   파이낸   청년취업아카데미   무료빅데이터  
데이터사이언스   KBS미디어청년취업아카데미   소프트웨어   컴퓨터보안프로그램   업가발

검색 키워드와 관련도가 높은 키워드입니다.

일간 검색량



28,900건

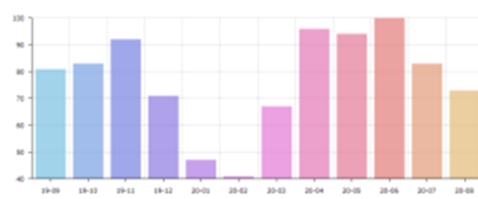


26,300건

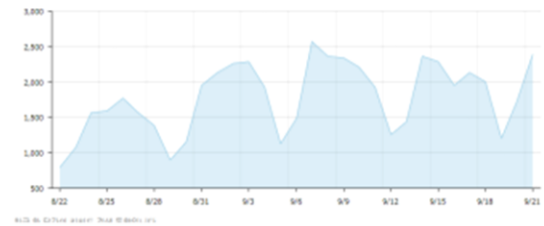
공통을 제외하고 한달간 키워드가 검색된 횟수입니다.

기간

최근 1년 검색량



일일 검색량



수집단위

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 1 [키워드 미리보기]

- 수집하기 전, 정보량 미리보기를 이용하여 확인할 수 있습니다.
- 수집할 키워드의 검색추이를 확인할 수 있습니다.
- ※ 네이버 채널의 키워드 정보를 제공하고 있습니다.
- ※ 미리보기에서는 연산자가 적용되지 않습니다.

2 미리보기 키워드를 입력 후 **[확인]**을 클릭하면 팝업창으로 확인 가능합니다.

포털/채널

뉴스

보유데이터

키워드 미리보기

확인

수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

수집키워드

1

키워드추가

키워드추가를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)

기간



2023-02-27

~



2023-03-06

1주

3개월

네이버 학술정보전체, 다음 웹문서, 트위터, 페이스북, 유튜브는 기간 설정이 불가능합니다.

수집단위

사용

사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 1 [수집키워드]

- 텍스트롬에는 **이용자가 키워드 란에 기입한 형태 그대로** 각 채널 검색창에 검색을 하여 그에 따른 검색 결과를 수집하게 됩니다.

※ 수집 전 대략적인 용량을 파악하고 싶다면 수집하고자 하는 채널에서 상세검색을 실시하여 기간을 설정하여 검색해본 후 나오는 결과 건 수를 확인해보세요.

### ※ 데이터 수집 범위



[텍스롬으로 분석한 사회 이슈] 주 52시간 근무제 2018.09.07.

빅데이터 일관처리 솔루션인 텍스트롬으로 짚어보았습니다. ▶ 수집 기간... 도구: 텍스트롬(Textom) ▶ 분석 도구: 텍스트롬(Textom), UCI-Net6, NetDraw ▶ 수집데이터량: 41...

빅데이터 일관처리 ... blog.naver.com/textom/221354509416

※ 검색 시 노출되는 게시글의 제목과 요약문을 수집합니다.

포털/채널

뉴스

보유데이터

키워드 미리보기

확인

수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색후이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

2

3

수집키워드

키워드추가

연산자 ▼

키워드추가를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)

기간

2023-02-27

~

2023-03-06

1주

3개월

1년

네이버 학술정보전체, 다음 웹문서, 트위터, 페이스북, 유튜브는 기간 설정이 불가능합니다.

수집단위

사용    사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 2 [키워드 추가]

- 여러 키워드를 동일한 설정(기간, 수집단위, 채널)으로 수집하고 싶을 때 사용하면 유용한 기능입니다.
- 분석의 목적과 주제에 맞는 정확한 데이터를 수집하기 위해서는 세밀한 키워드 선정이 필요합니다.

## 3 [연산자 사용]

- 네이버와 구글은 연산자 기능이 사용 가능한 채널입니다.
- ※ 상황에 맞게 적절한 특수문자(연산자)를 사용하면 더욱 정확한 데이터를 수집할 수 있습니다.



포털/채널

뉴스

보유데이터

키워드 미리보기

확인

수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

수집키워드

키워드추가

키워드추가를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)

1

기간

2023-02-27

~

2023-03-06

1주

3개월

네이버 학술정보전체, 다음 웹문서, 트위터, 페이스북, 유튜브는 기간 설정이 불가능합니다.

수집단위

사용

사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 1 [수집기간 설정]

- 설정 기간 내에 만들어진 데이터를 수집하게 됩니다.

※ 포털/SNS의 네이버 학술정보전체, 다음 웹문서, 유튜브, 트위터, 페이스북은 기간 설정이 적용되지 않으며, 수집이 진행되는 시점의 해당 채널 검색 결과를 그대로 수집하게 됩니다.

※ 뉴스의 언론사 전체 채널은 최대 3개월까지 수집 가능합니다.

- 긴 기간을 수집하려면 요청채널 서비스(유료)를 이용해주시거나 3개월 씩 나누어 여러 번 수집해 주셔야 합니다.

포털/채널

뉴스

보유데이터

키워드 미리보기

확인

수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.  
네이버 채널의 키워드 정보를 제공하고 있습니다.

수집키워드

키워드추가

키워드추가를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)

기간



2023-02-27

~



2023-03-06

1주

3개월

네이버 학술정보전체, 다음 웹문서, 트위터, 페이스북, 유튜브는 기간 설정이 불가능합니다.

2 수집단위

사용

사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

수집리스트생성 >>

## 2 [수집단위 설정]

- 수집단위는 일/주/월/년의 시간 단위 중 선택된 단위로 데이터를 최대 1,000건까지 수집하는 기능입니다.

※ 1년(365일)의 데이터를 일단위로 수집할 경우

- 최대 365,000건의 데이터를 수집할 수 있습니다.

- 동일한 기간을 수집단위 사용안함 으로 수집할 경우, 최대 1,000건 수집됩니다

※ 수집단위 기능을 사용할 수 있는 채널

- 네이버 블로그, 카페, 지식IN, 뉴스 / 다음 블로그, 카페, 뉴스 / 구글 뉴스입니다.

포털/채널

뉴스

보유데이터

채널별로 최대 1,000건의 문서를 수집합니다.

1

채널

수집정보

NAVER

네이버 전체

블로그

뉴스

카페

지식IN

학술정보전체

웹문서

Daum

다음 전체

티스토리

뉴스

카페

웹문서

\* 현재 Daum의 블로그 데이터의 경우, 티스토리 통합으로 인해 NAVER 게시글을 다수 제공하고 있어 데이터가 중복 수집 됩니다.

\* 중복 데이터 수집으로 발생한 불필요한 용량결제를 막고자 티스토리 수집만 지원합니다.

Google

구글 전체

뉴스

구글페이스북

웹문서

구글페이스북은 구글에서 수집하는 페이스북 문서입니다.

YouTube

유튜브

twitter

트위터

트위터, 페이스북은 보안정책으로 인해 일시적으로 수집이 원활하지 않을 수 있습니다.

2

수집리스트생성 >>

1 수집 가능한 채널은 해당 이미지와 같습니다.

- 네이버 : 전체/블로그/뉴스/카페/지식IN/학술정보전체/웹문서

- 다음 : 전체/티스토리/뉴스/카페/웹문서

- 구글 : 전체/뉴스/구글페이스북/웹문서

- 유튜브 / 트위터

2 조건 설정 후 [수집리스트 생성] 클릭 시 수집리스트가 생성됩니다.

## ☰ 데이터수집

포털/채널
뉴스
보유데이터

**1 키워드**  키워드추가

키워드추가를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)  
뉴스 채널은 원문수집만 가능합니다.

---

**2 기간** 2023-02-27 ~ 2023-03-06 1주 3개월

최대 3개월까지 수집 가능합니다.

---

**3 채널** 뉴스 채널은 수집단위를 사용할 수 없습니다. (채널별수집정보 : 제목, 본문, URL, 날짜)

뉴스 채널은 수집단위를 사용할 수 없습니다. (채널별수집정보 : 제목, 본문, URL, 날짜)

<input type="checkbox"/> KBS	<input type="checkbox"/> MBC	<input type="checkbox"/> SBS	<input type="checkbox"/> YTN
<input type="checkbox"/> 조선일보	<input type="checkbox"/> 동아일보	<input type="checkbox"/> 한겨레	<input type="checkbox"/> 경향신문
<input type="checkbox"/> 한국일보	<input type="checkbox"/> 서울신문	<input type="checkbox"/> 연합뉴스	<input type="checkbox"/> NEWS1

4 수집리스트생성 >>

**1 [키워드]** 란에 기입한 형태 그대로, 각 채널 검색창에 검색을 하고 그 검색 결과를 수집하게 됩니다.

**2** 원하는 수집 **[기간]**을 설정할 수 있으며 설정 기간 내에 만들어진 데이터를 수집하게 됩니다.

**3** 수집 가능한 채널은 이미지(총 12개 언론사)와 같습니다.

**4** 조건 설정 후 **[수집리스트 설정]** 클릭 시 수집리스트가 생성됩니다.

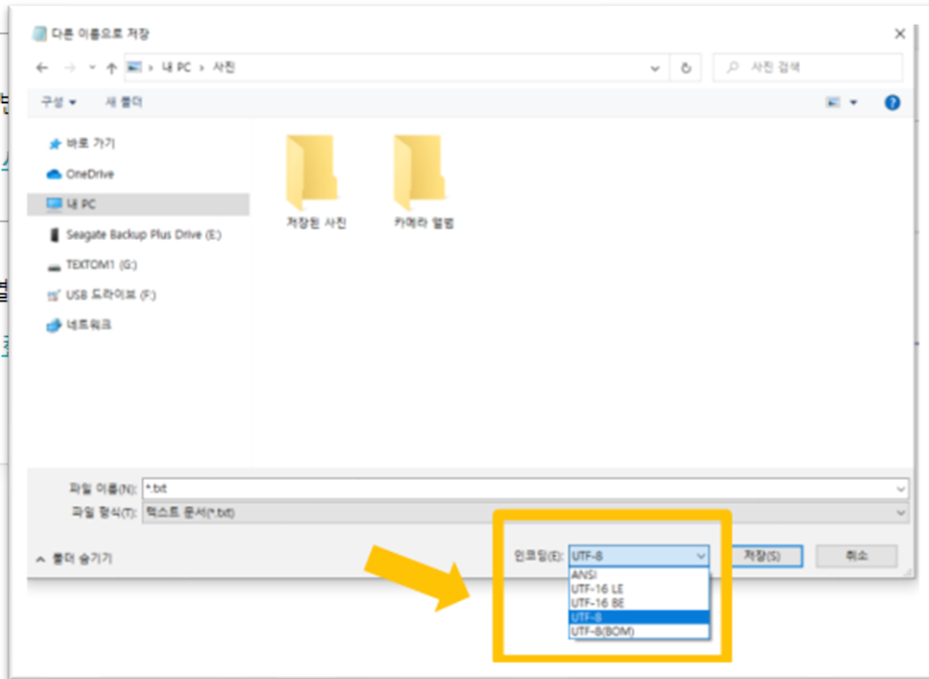
포털/채널

뉴스

보유데이터

### 1 데이터업로드

txt, pdf, xls, xlsx 파일만 업로드 가능합니다.  
txt : UTF-8 인코딩 파일로 업로드합니다.



1 가지고 있는 데이터를 텍스트에 업로드하여 정제, 분석, 시각화 할 수 있습니다.

※ TXT, PDF, XLSX, XLS 파일만 업로드 가능합니다.

※ TXT 파일은 UTF-8 인코딩 파일로 업로드 해 주셔야 합니다.

※ PDF 파일은 스캔본 및 이미지 파일은 이미지로 인식되어 분석되지 않습니다.

포털/채널

뉴스

보유데이터

데이터업로드

txt, pdf, xls,xlsx 파일만 업로드 가능합니다.

txt : UTF-8 인코딩 파일로 업로드합니다.

2 시트지정

첫 번째  두 번째  세 번째

직접지정 :

적용 할 시트를 지정합니다. 직접지정 : 예) 4,5,6

3 컬럼지정

A열  B열  C열  D열  E열  F열

직접지정 :

적용 할 컬럼을 지정합니다. 직접지정 : 예) D,G,H

4

수집리스트생성 >>

## 2 [시트설정]

- 엑셀파일을 업로드한 경우 파일 내 분석에 사용할 시트를 지정할 수 있습니다.

## 3 [컬럼지정]

- 엑셀 파일 업로드 시, 분석에 사용할 열을 지정할 수 있습니다.

※ 단, 컬럼 다중 선택 시 컬럼들은 한 컬럼으로 합쳐져 분석되게 됩니다.

4 조건 설정 후 [수집리스트 생성] 클릭 시 수집리스트가 생성됩니다.

1

포털/채널    뉴스    보유데이터

총 151 건    수집이 완료된 데이터는 30일 전까지만 삭제됩니다.

키워드검색    검색 결과 151 / 건    삭제

키워드	채널	기간	수집날짜	용량	수집상태
<input type="checkbox"/> "마우스"-키보드	네이버	2023.03.13 ~2023.03.20	2023.03.20	32.55 KB	대기 75번
<input type="checkbox"/> 무신사 +뉴진스 -LG그램	네이버(블로그,카페,뉴스)	2022.10.01 ~2023.02.12	2023.02.13	235.89 KB	
<input type="checkbox"/> 메가팩스 메가스터디 +뉴진스	네이버(블로그,카페,뉴스)	2022.10.01 ~2023.02.12	2023.02.13	40.38 KB	
<input type="checkbox"/> 누데이크 +뉴진스	네이버(블로그,카페,뉴스)	2022.12.01 ~2023.02.12	2023.02.13	580.79 KB	
<input type="checkbox"/> LG그램 스타일 빅버니 +뉴진스	네이버(블로그,카페,뉴스)	2022.12.01 ~2023.02.12	2023.02.13	155.01 KB	
<input type="checkbox"/> 누데이크 -뉴진스	네이버(블로그,카페,뉴스)	2019.01.01 ~2023.02.12	2023.02.13	8.09 MB	
<input type="checkbox"/> 메타버스 +선글라스	네이버(블로그,카페,뉴스)	2020.01.01 ~2022.12.31	2023.01.11	711.21 KB	
<input type="checkbox"/> "스마트글라스"	네이버(블로그,카페,뉴스)	2020.01.01 ~2022.12.31	2023.01.11	2.27 MB	
<input type="checkbox"/> "스마트글래스"	네이버(블로그,카페,뉴스)	2020.01.01 ~2022.12.31	2023.01.10	1.38 MB	
<input type="checkbox"/> 메타버스 +안경	네이버(블로그,카페,뉴스)	2020.01.01 ~2022.12.31	2023.01.10	3.15 MB	

5

데이터 미리보기

채널    섹션    수집량(건)    용량

조화할 데이터를 선택해주세요.

채널별 수집량 시각화

데이터 전처리하기 >>

(예상 차감 용량: 0.00 KB)

잔여 용량을 확인해주세요!

잔여 용량: 11.91 GB    용량 추가

※ 수집량이 1,000건인 데이터가 많습니다. 왜 그런가요?  
 오픈API를 통해서 데이터를 수집하는 경우, 섹션별로 한번에 최대 1,000건까지 수집할 수 있습니다  
 1,000건 이상의 데이터를 수집 해야 하는 경우 [수집단위 기능](#)을 사용해주세요

- 1 수집한 채널별로 수집 리스트를 확인할 수 있습니다.
- 2 키워드를 통해 검색이 가능합니다.
- 3 수집한 데이터를 확인할 수 있는 구역입니다.
- 4 데이터의 체크박스를 체크 후 해당 버튼을 누르면 수집리스트에서 삭제가 가능합니다.
- 5 좌측 수집 데이터 구역에서 데이터를 클릭하면 미리보기 영역에 해당 데이터의 채널, 섹션, 수집량 등의 미리보기 정보를 제공합니다.

포털/채널

뉴스

보유데이터

2 / 2

수집완료 리스트의 데이터는 30일이 지나면 삭제됩니다.

<input type="checkbox"/>	키워드	채널	기간
<input type="checkbox"/>	갤럭시	네이버(블로그)	2022.06.23 ~ 2022.06.30
<input type="checkbox"/>	<b>1</b> 폴더 아이콘 나이키	네이버(블로그,카페,뉴스)	2022.01.29 ~ 2022.01.07
<input type="checkbox"/>	2022-01-29 ~ 2022-02-04		
<input type="checkbox"/>	2022-01-22 ~ 2022-01-28		
<input type="checkbox"/>	2022-01-15 ~ 2022-01-21		
<input type="checkbox"/>	2022-01-08 ~ 2022-01-14		
<input type="checkbox"/>	2022-01-01 ~ 2022-01-07		

- ① 수집 시, [수집단위] 기능을 사용한 데이터는 키워드 앞에 [폴더 아이콘](#)이 있으며 해당 리스트를 클릭하면 수집 시 선택한 수집단위로 나뉘어진 하위 리스트가 다음과 같이 펼쳐집니다.

※ 수집단위 기능

- 일/주/월/년의 시간 단위 중 선택된 단위로 데이터를 최대 1,000건 수집하는 기능입니다.
- 예를 들어, 1년(365일)의 데이터를 일단위 수집할 경우, 최대 365,000건의 데이터를 수집할 수 있습니다. 동일 기간을 수집단위 사용안함으로 수집할 경우, 최대 1,000건 수집됩니다.



## 데이터 미리보기

▶ AI교육

2022.01.01 ~ 2022.01.01

1

채널	섹션	수집량(건)	용량
네이버 <b>NAVER</b>	블로그	81099	2 31.64 KB

### 원문 데이터 보기

데이터명	생성날짜	용량
AI교육	2023-01-09	63.54 MB

50 / 166351

AI도 대체할 수 없는 직업, 전기산업기사 및 전기기사 전망! & 오늘의 주제는 바로 AI가 대체할 수 없는 유망직종인데요. 바로 전기기사와... 합격으로 가는 발걸음, '독학'이 아닌 철저하게 계획된 '체계적인 교육' 속에서..  
<https://blog.naver.com/durudong/222871562391>  
 naver | blog

[캠프후기] 영상미디어 제작 Day-4  
 디자인씽킹 #퍼실리테이터 #비즈니스클 #취창업역량강화 #모의창업 #리더십교육 #인성교육 #AI인적성검사 #AI모의연결 #전문강사양성 #동아리프로그램 #0  
<https://blog.naver.com/fedulab2050/222871147249>  
 naver | blog

[2204 Korea] 노원 불빛 정원  
 \* 압턴 기자 7-8랑 정도에 걸쳐서 내부에 시간에 관한 여러가지 전사를 해두었는데 아이들 교육에도 좋을듯 한 전사가 많았다. AI를 이용해서 40년후의 모습을  
<https://blog.naver.com/coolhs99/222871601358>  
 naver | blog

서울대, 연고대도 옛말-신설학과, 계약학과 전공, 학과선택이...  
 정확한 전략으로 AI 대가의 조연- "소용민처럼 교육하라" "내일 뉴스에 나올 이야기보다 미래를 준비해야 한다" 처럼 미래는 스마트하게 준비된 자의 것! 뉴스  
<https://blog.naver.com/ahuhak/222871161940>  
 naver | blog

세마역 아파트 구조 오피스텔 청광플러스원 분양정보  
 온암플 스마트시티 개발은 7000억원 투자규모의 대형 프로젝트로 오산시청 동측... 쇼핑 시설이 어우러진 복합상업시설, 문화·교육시설, 명품 주거단지가 연  
[https://blog.naver.com/modelhouse\\_s/222870942292](https://blog.naver.com/modelhouse_s/222870942292)  
 naver | blog

광장르엣지29 광진구 광장동 주상복합 아파트 공급 소식  
 광장르엣지29는 교육 관련 시설이 좋은 학세권에 있기도 합니다. 양진 학교과... AI IoT 시스템을 도입해서 음성인식으로 주택 제어가 가능하네요. 여기에 무인  
<https://blog.naver.com/ajt5144/222868343414>  
 naver | blog

[8월호] 센터 이슈 N 소식  
 총주상고의 주요 취업 지원 프로그램으로는 자기소개서-이력서 쓰기 특강, 면접특강, 산업체 탐방, 동행 면접, 위탁 교육 등이 있습니다. 또, 최근 트렌드인 AI화  
[https://blog.naver.com/moe\\_centralwork/222862250686](https://blog.naver.com/moe_centralwork/222862250686)  
 naver | blog

### ① [리스트별 수집 내역 확인]

- 데이터리스트를 선택하면 우측 [데이터 미리보기] 대시보드에서 [채널별/섹션별] 수집된 건 수와 용량 수치를 확인 할 수 있습니다.

### ② [원문데이터 미리보기]

- 각 용량을 클릭하면 수집데이터의 원문을 팝업창으로 미리 살펴볼 수 있습니다.

## 데이터 미리보기

▶ AI교육

2022.01.01 ~ 2022.01.01

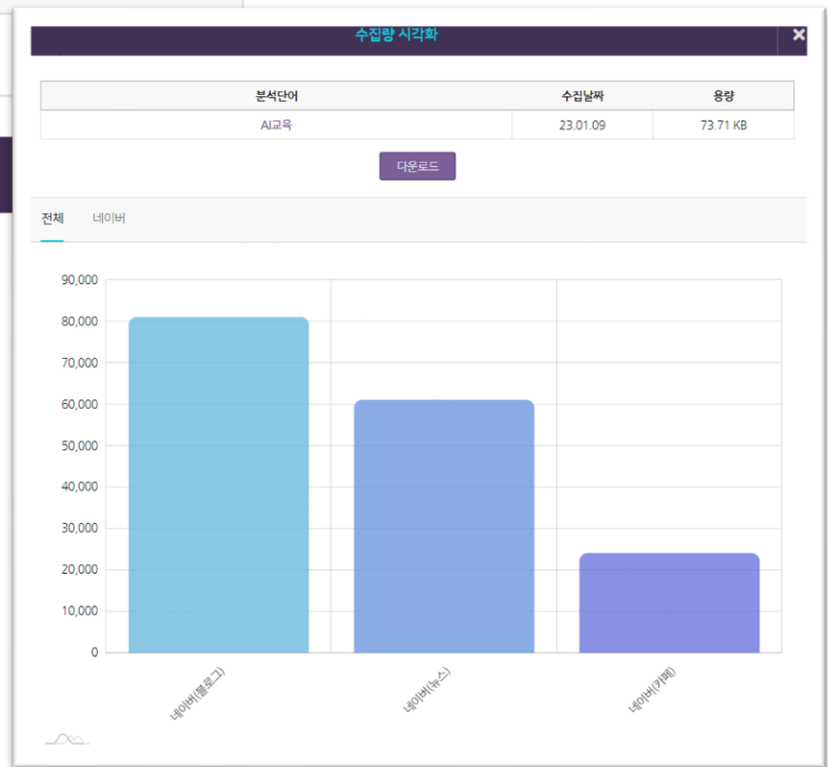
채널	섹션	수집량(건)	용량
네이버 NAVER N	블로그	81099	31.64 KB
	뉴스	61148	26.49 KB
	카페	24104	15.59 KB

1

채널별 수집량 시각화

2

데이터 전처리하기»



### ① [채널별 수집량 시각화]

- 각 채널별로 수집된 데이터 양을 확인 할 수 있는 팝업창이 등장합니다.
- 팝업창 내 **[다운로드]** 버튼을 클릭하면 그래프가 png 파일로 다운로드 됩니다.

### ② 데이터를 확인하였다면 해당 수집리스트를 체크한 후 [데이터 전처리하기]

버튼을 클릭하여 전처리 과정을 진행하실 수 있습니다.

1

데이터 정제 선택    형태소 분석 선택

분리정제    중복제거    필터링    Window-Size

[분리 정제] 도구를 이용하면 제목과 내용을 분리하거나 통합하여 데이터를 분석할 수 있습니다.  
[분리 정제]에서 선택된 정제 방식에 따라 [중복 제거] 및 [필터링] 기능은 제한될 수 있습니다.

정제 방식     전체(제목+본문)     제목     본문

적용 >>

2

데이터 미리보기

▶ 안동시  
2022-09-02 ~ 2022-12-01

정제이력    삭제

기능명	상세내역
-----	------

데이터 전처리하기 >>

① 데이터 전처리에서는 데이터 정제, 형태소 분석을 진행할 수 있습니다.

※ 데이터 정제

- 중복제거, 키워드 필터링, window-size 등의 기능을 활용하여 원문 데이터를 분석이 가능한 형태로 가공하는 과정

※ 형태소 분석

- 단어를 구성하는 각각의 형태소들을 인식하고 용언의 활용, 불규칙 활용이나 축약, 탈락현상이 일어난 형태소를 원형으로 복원 하는 과정

② [정제이력]은 데이터 전처리에서 적용한 정제, 형태소 분석 과정을 저장 및 삭제할 수 있는 영역입니다.

데이터 정제 선택

형태소 분석 선택

분리정제 중복제거 필터링 Window-Size

1

[분리 정제] 도구를 이용하면 제목과 내용을 분리하거나 통합하여 데이터를 분석할 수 있습니다.  
[분리 정제]에서 선택된 정제 방식에 따라 [중복 제거] 및 [필터링] 기능은 제한될 수 있습니다.

정제 방식  전체(제목+본문)  제목  본문

적용 >>

반드시 선택해야하는 기능

## 1 [분리정제]

- 제목과 내용을 분리하거나 통합하여 분석할 수 있습니다.
- 분리 정제에서 선택된 항목에 대해 중복 제거, 필터링 기능이 적용됩니다.
- **반드시 진행해야하는 단계**로 선택 후 **[적용]** 버튼을 꼭 클릭하세요.

※ [전체] 선택 시

- 전체(제목과 본문)을 통합하여 분석을 진행하며, 중복 제거나 필터링 기능이 제목과 본문에 모두 적용됩니다.

※ [제목] 선택 시

- 제목만 분석을 진행하며, 중복 제거나 필터링 기능이 제목에만 적용됩니다.

※ [본문] 선택 시

- 본문만 분석을 진행하며, 중복 제거나 필터링 기능이 본문에만 적용됩니다.

데이터 정제 선택

형태소 분석 선택

분리정제   **중복제거**   필터링   Window-Size

**1** [중복 제거] 도구를 이용하면 제목이나 내용, URL이 중복되는 데이터를 제거합니다.  
중복을 제거할 컬럼을 선택하여 조건을 추가해 주세요.

컬럼 선택    제목    내용    URL

적용 >>

선택사항

정제이력

삭제

<input type="checkbox"/>	기능명	상세내역
<input type="checkbox"/>	중복 제거	제목

<단일 컬럼 선택>

정제이력

삭제

<input type="checkbox"/>	기능명	상세내역
<input type="checkbox"/>	중복 제거	제목, 내용

<복수 컬럼 선택>

## ① [중복제거]

- 중복되는 데이터를 [제목, 내용, URL] 기반으로 제거합니다.
- 복수 조건 등록 시 두 컬럼 모두 동일한 데이터를 모두 삭제하는 것이 아니라, 각각의 컬럼에서 동일한 데이터를 삭제합니다.
- 선택한 항목이 있다면 **[적용]** 버튼을 꼭 눌러 정제이력을 확인해주세요.

### ※ 단일 컬럼 선택 시

- 제목 컬럼 선택 시 제목 컬럼에서 동일한 데이터만 삭제합니다.

### ※ 복수 컬럼 선택 시

- 제목, 내용 컬럼 선택 시 제목 컬럼에서 중복 데이터 삭제, 내용 컬럼에서 동일한 데이터 삭제를 각각 수행합니다. (제목과 내용이 함께 동일한 데이터를 삭제하지는 않습니다.)

데이터 정제 선택
형태소 분석 선택

분리정제
중복제거
필터링
Window-Size

[필터링] 도구를 이용하면 제목이나 내용에 특정 단어의 포함 여부에 따라 데이터를 추출할 수 있습니다.  
[필터 추가하기] 버튼을 눌러 필터링 조건을 생성해주세요.

1

필터링추가

2

컬럼 선택

제목
내용

필터 유형

AND 조건     OR 조건

텍스트 필터

포함
제외

+

포함
제외

-

컬럼 선택

제목
내용

필터 유형

AND 조건     OR 조건

텍스트 필터

포함
제외

+

적용 >>

선택사항

① [필터링 추가] 버튼을 누르면 아래에 텍스트 필터링 설정 란이 생성됩니다.  
필터링 조건은 최대 2개까지 설정 가능합니다.

② 필터링 조건을 적용할 컬럼(제목/내용)을 선택합니다.

- [분리정제] 단계에서 선택한 정제 방식을 기반으로 컬럼을 선택 하셔야 합니다.

※ 전체(제목+내용) 선택 : 제목, 내용 컬럼 모두 선택가능

※ 제목 선택 : 제목 컬럼만 선택(내용 컬럼 선택 불가)

※ 내용 선택 : 내용 컬럼만 선택(제목 컬럼 선택 불가)

데이터 정제 선택

형태소 분석 선택

분리정제 중복제거 **필터링** Window-Size

[필터링] 도구를 이용하면 제목이나 내용에 특정 단어의 포함 여부에 따라 데이터를 추출할 수 있습니다.  
[필터 추가하기] 버튼을 눌러 필터링 조건을 생성해주세요.

1

필터링추가

컬럼 선택

제목

내용

조건 제거

2

필터 유형



AND 조건



OR 조건

텍스트 필터

포함

제외

팬더

+

포함

제외

너구리

-

컬럼 선택

제목

내용

조건 제거

필터 유형



AND 조건



OR 조건

텍스트 필터

포함

제외

사자

+

적용 >>

선택사항

③ 필터 유형을 선택합니다.

※ AND 조건 : AND 조건일 경우 제목에 팬더와 너구리를 포함한 결과를 생성합니다.

※ OR 조건 : OR 조건일 경우 제목에 팬더 또는 너구리를 포함한 결과를 생성합니다.

데이터 정제 선택
형태소 분석 선택

분리정제
중복제거
필터링
Window-Size

[필터링] 도구를 이용하면 제목이나 내용에 특정 단어의 포함 여부에 따라 데이터를 추출할 수 있습니다.  
[필터 추가하기] 버튼을 눌러 필터링 조건을 생성해주세요.

필터링추가

컬럼 선택 제목 내용

필터 유형  AND 조건  OR 조건

1 **텍스트 필터**

포함
제외

+

포함
제외

-

조건 제거

컬럼 선택 제목 내용

필터 유형  AND 조건  OR 조건

**2** **텍스트 필터**

포함
제외

+

조건 제거

적용 >>

선택사항

## 1 [텍스트 필터]

- 필터링할 조건(유형)을 선택 후 필터링할 텍스트를 입력합니다.
- (+), (-) 클릭 시 텍스트를 추가 및 삭제할 수 있습니다.
- ※ 포함: 입력한 키워드를 포함하는 내용들을 필터링합니다.
- ※ 제외: 입력한 키워드를 제외한 나머지 내용들을 필터링합니다.

## 2 해당 조건으로 두 개의 필터링을 선택하고 [적용] 버튼을 클릭할 경우 아래 선택한 필터링 조건을 모두 적용합니다.

- ※ (필터링1) 제목에서 팬더와 너구리를 포함하고, (필터링2) 내용에서 사자를 제외한 데이터



데이터 정제 선택

형태소 분석 선택

분리정제   중복제거   필터링   **Window-Size**

[Windows-Size] 도구는 분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상범위를 지정합니다. 주제어와 밀접한 단어들에 대한 데이터만 제공합니다.

1

키워드

사이즈

3개



적용 >>

선택사항

## 1 [Window-Size]

- 한 문서 내에서 입력한 키워드를 기준으로 앞뒤로 단어 개수 범위를 지정하여 데이터를 정제할 수 있습니다.
- 선택 후 **[적용]** 버튼을 꼭 눌러 정제이력 리스트를 확인하세요.

※ 예시) “나랏말싸미 동국에 달아 문자와로 서르 사맛디 아니할씨” 라는 문장에서 키워드를 [달아] / 사이즈를 2로 설정한다면, [달아]의 앞뒤로 2개까지의 단어인 “나랏말싸미/동국에/달아/문자와로/서르”까지 추출된 결과를 확인하게 됩니다.

데이터 정제 선택
형태소 분석 선택

---

1 분석언어
 

한국어
영어

---

2 분석기 ?

Espresso K
MeCab

Espresso K는 고유명사, 복합명사를 그대로 결과값에 반영하며  
 MeCab은 띄어쓰기와 상관없이 사전을 참조하여 어휘를 구분합니다.  
 예) "사회복지학과"를 Espresso K는 "사회복지학"으로 Mecab은 "사회 복지 학과"로 정제합니다.

---

분석품사

단순품사

상세품사

? 품사태그
 

초기화
결과 미리보기

▶ 체언

일반 명사(NNG)
 고유명사(NNP)
 의존명사(NNB)
 단위명사(NNBC)
 수사(NR)
 대명사(NP)

---

▶ 용언

▶ 어근

동사(VV)
 형용사(VA)
 어근(XR)

---

▶ 수식언

▶ 독립언

관형사(MM)
 일반부사(MAG)
 접속부사(MAJ)
 감탄사(IC)

---

▶ 접미사

▶ 접두사

명사접미사(XSN)
 동사접미사(XSV)
 체언접두사(XPN)

적용 >>

형태소 분석은 단어를 구성하는 각각의 형태소들을 인식하고 용언의 활용, 불규칙 활용이나 축약, 탈락현상이 일어난 형태소를 원형으로 복원 하는 과정을 의미하며, 형태소 분석기는 텍스트를 형태소 단위로 분석하고 품사를 함께 출력해주거나 특정 품사에 해당하는 형태소만 선별해 주는 패키지를 의미합니다.

**1** 분석할 언어를 선택합니다.

**2** 각 분석기의 특성을 참고하여 형태소 분석기를 선택해줍니다.

※ Espresso K는 고유명사, 복합명사를 그대로 결과값에 반영하며

※ MeCab은 띄어쓰기와 상관없이 사전을 참조하여 어휘를 구분합니다.

예) "사회복지학과"를 Espresso K는 "사회복지학"으로 Mecab은 "사회 복지 학과"로 정제합니다.

데이터 정제 선택

형태소 분석 선택

MeCab은 띄어쓰기와 상관없이 사전을 참조하여 어휘를 구분합니다.  
 예) "사회복지학과"를 Espresso K는 "사회복지학"으로 Mecab은 "사회 복지 학과"로 정제합니다.

분석품사

1

단순품사

상세품사

품사 태그

2

초기화

결과 미리보기

- ▶ 체언
  - 일반 명사(NNG)  고유명사(NNP)  의
- ▶ 용언
  - 동사(VV)  형용사(VA)
- ▶ 수식언
  - 관형사(MM)  일반부사(MAG)  접속
- ▶ 접미사
  - 명사접미사(XSN)  동사접미사(XSV)
  - 형용사접미사(XSA)
- ▶ 한글 이외
  - 외국어(SL)  숫자(SN)

사용자사전 ?

사용

사용안함

형태소 분석 결과 미리보기 (예시문장)

분석 품사 선택

- ▶ 체언
  - 일반 명사(NNG)  고유명사(NNP)  의존명사(AN)  단위명사(NE)  수사(NR)  대명사(PO)
- ▶ 용언
  - 동사(VV)  형용사(VA)
- ▶ 수식언
  - 관형사(MM)  일반부사(MAG)  접속부사(MAJ)
- ▶ 접미사
  - 명사접미사(XSN)  동사접미사(XSV)
  - 형용사접미사(XSA)
- ▶ 한글 이외
  - 외국어(SL)  숫자(SN)

예시문장	형태소 분석 결과
서울역에서 시를 읽다가는 지하철 1호선 한 구간의 개찰입니다.	서울역(NNG)+에서(KE)+는(EO)+지하철(NG)+1(ISO)+호선(NNEC)+한(NNG)+구간(NG)+의(EO)+개찰(NG)+입니다(VCP)
아파트의 중간층에는 국토 1인 승객만이 있지만 선택할 때도 화일로 올라 가깝은 재건축을 위해 가행자를 설치 하는 상태입니다.	아파트(NG)+의(EO)+중간층(NG)+에는(EO)+국토(NG)+1(ISO)+인(NG)+승객(NG)+만이(PO)+있지만(VC)+선택(NG)+할(NG)+때(KE)+도(EO)+화일로(NG)+올라(NG)+가(NG)+깝은(NG)+재건축(NG)+을(KE)+위해(EO)+가행자(NG)+를(KE)+설치(NG)+하는(NG)+상태(NG)+입니다(VCP)
그러면 함께 서울역에서 시를 읽까지 유용하게 할어 할까요? LET'S GO!	그러면(MAJ)+함께(MAG)+서울역(NNG)+에서(KE)+는(EO)+시(NG)+를(NG)+읽(NG)+까지(EO)+유용(NG)+하게(EO)+할(NG)+어(NG)+할(NG)+까요(NG)+?+LET'S(NG)+GO(NG)+!+입니다(VCP)
최근 이곳에 서울 4대문 중 하나인 숙정문입니다. 우측을 보시면 처마 앞 쪽에는 소나무의 가지가 고를수 있습니다.	최근(NG)+이곳(NG)+에(EO)+서울(NG)+4(ISO)+대문(NG)+중(NG)+하나(NG)+인(NG)+숙정문(NG)+입니다(VCP)+.우측(NG)+을(KE)+보(NG)+시면(VC)+처마(NG)+앞(NG)+쪽(NG)+에는(EO)+소나무(NG)+의(EO)+가지(NG)+가(NG)+고(NG)+를(NG)+수(NG)+있(NG)+습(NG)+다(VCP)

선택 품사 적용

1 분석 품사를 선택할 수 있습니다.

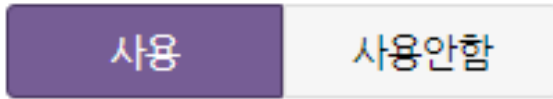
※ 단순품사: 명사, 형용사, 동사, 외국어, 숫자를 선택할 수 있습니다.

※ 상세품사: 체언, 용언, 어근, 수식언, 독립언, 접미사, 접속부사, 외국어, 숫자를 선택할 수 있습니다.

2 분석기별 품사 태그를 확인할 수 있습니다.

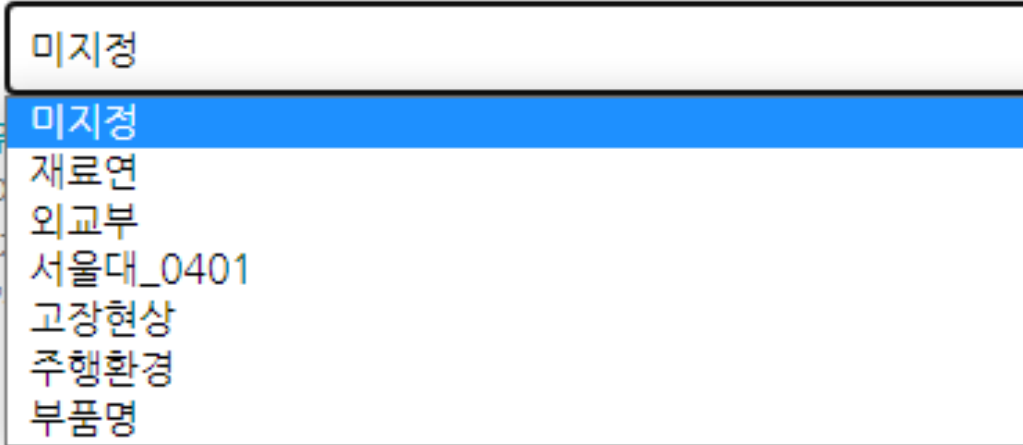
3 선택한 품사가 예시문장에 바로 적용되어 정제데이터에 선별될 단어를 미리 확인할 수 있습니다.

## 1 사용자사전 ?

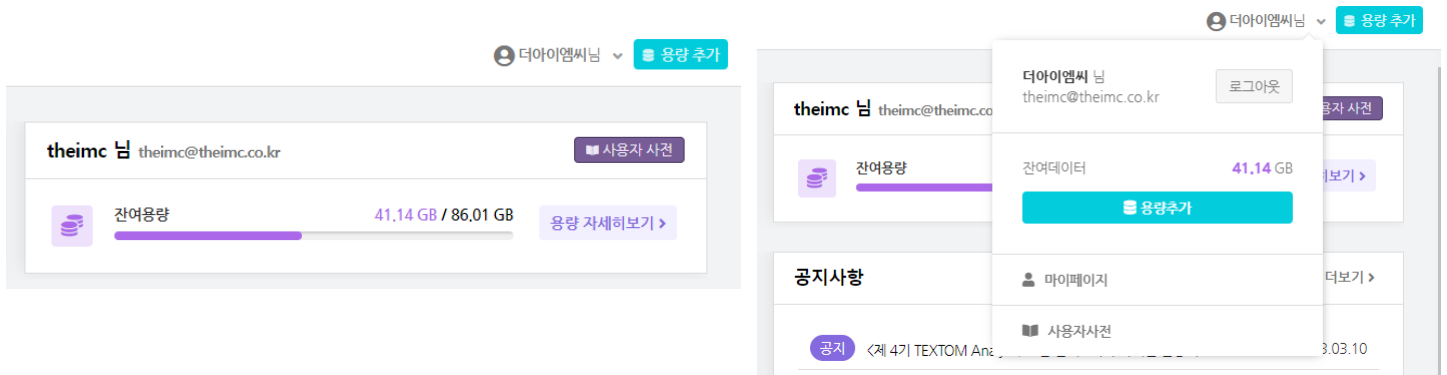


## 2 그룹지정

정제할 키워드  
텍스트를 이  
인 분석이  
예)'워라벨'



## 3



### 1 [사용자 사전]

- 유사한 주제의 데이터 또는 동일한 주제의 기간 및 채널만 다른 데이터를 반복 정제해야 하는 경우, 변경할 단어 들을 사전으로 구축해두면, 형태소 분석 과정에서 지정해둔 수정 단어들로 일괄 변경되는 기능입니다.

2 사전은 여러 가지를 설정해둘 수 있으며 미리 설정해둔 사용자 사전을 [그룹지정]을 통해 불러와 사용할 수 있습니다.

3 사용자 사전은 대시보드 우측 및 내 정보 확인 구간에서 [사용자 사전] 버튼 클릭 시 편집 페이지로 이동할 수 있습니다.

The screenshot shows a web interface for keyword refinement. At the top, a table displays data for '0320 테스트' (0320 Test) with a creation date of 2023-03-20 and a size of 35.52 KB. Below this is a search bar where '키워드 검색' (Keyword Search) is active, with checkboxes for '정확한 일치' (Exact match) and '대소문자 구분' (Case sensitive). A search input field contains '변경할 단어' (Word to be changed). To the right, there are fields for '수정할 단어' (Word to be corrected) and a '변경' (Change) button. The main content area is divided into two columns: 'TF' (Term Frequency) and 'N-gram'. The TF column lists terms like '안동', '안동시', '안동시청' with their respective frequencies. The N-gram column lists terms like '안동', '안동시', '안동시청' with their respective counts. On the right side, there are two panels: '추천 단어' (Recommended words) and '수정 내역' (Correction history). The '추천 단어' panel shows terms like '오카동', '내년 오전', '내년 오전', '센트럴 자이', and '센트럴 자이'. The '수정 내역' panel shows terms like '브행고', '브행고 (추진/대소구분)', 'K POP', 'KPOP (사용자/대소구분)', and 'k pop'. At the bottom, there are two buttons: '변경한 내용 적용' (Apply changed content) and '사용자 사전' (User dictionary).

- 1 데이터에 대한 간략한 정보가 표시됩니다.
- 2 원문 데이터, 정제 데이터, N-gram 결과에 대해 키워드 검색을 할 수 있습니다.
- 3 변경할 단어, 수정할 단어를 입력해 키워드 변경을 할 수 있습니다.
- 4 TF(단어 빈도), N-Gram 결과를 확인할 수 있습니다.
- 5 정제 추천 단어를 확인하고, 정제에 적용할 수 있습니다.
- 6 수정 내역을 확인할 수 있습니다.
- 7 정제하고 있는 수정 내역을 사용자 사전에 등록하거나, 기존 보유하고 있던 사용자 사전을 불러올 수 있습니다.
- 8 변경한 내용 적용을 클릭하면 키워드 정제 내역이 적용됩니다.

1  정확한일치  대소문자구분

키워드 검색

2 원문  🔍

3 원문 검색 결과 8165 건의 데이터가 존재합니다.

갤럭시s21 플러스 색상 후속모델 비교 갤럭시s21 플러스 색상부터 시작해서 다양한 스펙을 알아보기로 한 것도 바로 이런... 갤럭시s21 플러스 색상을 보자면 팬텀 실버, 팬텀 블랙, 팬텀 바이올렛 그리고 팬텀..  
 갤럭시 3in1 무선충전기, 우승테크 WL-31 Qi 인증 고속 멀티... 제품은 갤럭시 전용과 아이폰 전용이 각각 출시가 이뤄졌으므로 처음 구매할 때 반드시 확인하고 구매해야 돼요 저는 갤럭시 스마트폰을 사용하다보니 나머지..  
 갤럭시s22 울트라 버건디 사려면 이 곳 추천 오늘 포스팅 통해 간단히 알려드릴테니 갤럭시s22 울트라 버건디 같은 최신폰 구매... 하지만 이 프로그램은 갤럭시s22 울트라 버건디를 48개월 할부로 구입한 다음에..  
 경기가 안좋은 요즘, 핸드폰 교체가 필요하다면 '갤럭시 Jump2... 제품을 추천드리게 된 계기 지금 구매하기 좋은 삼성 갤럭시 Jump2 경기가 너무... 추천 드리는 갤럭시 Jump2는 이런 부모님들께서 구매하시기 굉장히 좋은 조건을 갖고..  
 케이스티파이 갤럭시 케이스! 지존 Zizon 첫구매... 갤럭시S20은 이제 구형이라 케이스도 이쁘게 안나온다고오오오 ππππ!!! 현대... 다양하고 갤럭시

## 1 [키워드 검색]

- [정확한일치], [대소문자구분] 여부를 설정하여 검색할 수 있습니다.

※ 정확한일치: 정확하게 일치하는 문자열만 검색합니다.

※ 예시) '이번v메르sv사태'와 '이번메르스사태'에서 '메르스'를 '코로나'로 변경하고자 할 때,  
 - 정확한일치는 띄어쓰기가 있는 '이번v메르sv사태'를 '이번v코로나v사태'로 변경하고, '이번메르스사태'는 변경하지 않습니다.  
 - 부분일치는 '이번v메르sv사태', '이번메르스사태'를 '이번v코로나v사태', '이번코로나v사태'로 '메르스'를 포함한 모든 문자열을 변경합니다.

※ 대소문자구분: 대문자 소문자 구분 여부를 지정합니다.

2 원문 데이터, 정제 데이터, N-gram 결과에 대해 키워드 검색을 할 수 있습니다.

※ 원문: 정제를 적용하기 전 원문 데이터입니다.

※ 정제: 데이터 전처리(데이터 정제, 형태소 분석)가 적용된 후의 데이터입니다.

※ N-Gram: N개의 연속적인 단어 나열로, 결과에서는 2-gram의 결과를 제공합니다.

3 검색 결과가 나타나는 영역입니다. 스크롤을 내리면 데이터를 더 확인할 수 있으며, 2번의 원문/정제 구간을 설정하면 아래의 내용 또한 같이 변동됩니다.

## 1

### 키워드 검색

정확한일치  대소문자구분

정제 검색 결과 4 건의 데이터가 존재합니다.

정제 검색 결과 4 건의 데이터가 존재합니다.

정제 검색 결과 4 건의 데이터가 존재합니다.

경상북도 안동 아파트 안동 센트럴 자이 매머 지역 실거 리스트 매머 전월세 전매 지역 오늘 실거 신고 경상북도 안동 문흥동 경상북도 안동 당북동 경상북도 안동 안기동 경상북도 안동 송현동  
 안동 황고집 숲들 구이 안동정 하동 맛길 경북 안동 강남 길 현진 에버 아파트 경북 안동 경허동 현진 에버 아파트 영업시간 휴무일 변동  
 안동 아파트 실거 신고 현황 동시 아파트 실거 신고 내역 단위 전용 면적 시군 계약 흑년 총 실거 금액 이전 최고 최근 전세 이전 최고 안동 센트럴 자이 경상북도 안동  
 안동 아파트 실거 건수 안동 센트럴 자이 안동 센트 합격 현황 안 동시 아파트 신고 내역 동시 아파트 실거 내역 동시 분양 실거 내역

### 키워드 검색

N-gram 검색 결과 9 건의 데이터가 존재합니다.

N-gram 검색 결과 9 건의 데이터가 존재합니다.

N-gram 검색 결과 9 건의 데이터가 존재합니다.

11 아파트 실거 4  
 29 동시 아파트 3  
 102 에버 아파트 2  
 116 안동 아파트 2  
 220아파트 안동 1  
 433 안동 아파트 1  
 530아파트 경북 1  
 609아파트 신고 1  
 740아파트 영업시간1

1 [키워드 검색] 기능은 데이터 정제 과정 중 데이터 확인, 연결된 단어 확인 적용된 결과 확인을 위해 활용할 수 있습니다.

#### ※ 원문, 정제 검색

- 키워드 정제를 진행하는 중 특정 키워드가 원문 데이터와 정제 데이터에서 어떤 식으로 활용되고 있는지 보고자 할 때 해당 기능을 활용할 수 있습니다.

원문, 정제 데이터에 '아파트'을 검색해 '아파트'이라는 단어가 원문과 정제 데이터에서 어떤 식으로 활용 되는지 확인할 수 있습니다.

#### ※ N-Gram 검색

- 특정 키워드가 어떤 단어와 함께 사용되고 있는지 파악하여 정제하는데 활용할 수 있습니다.

N-Gram에 '아파트'을 검색해 '아파트'이라는 단어가 어떤 단어들과 많이 쓰였는지 볼 수 있습니다. 해당 결과를 참고해 손쉽게 키워드 정제를 할 수 있습니다.

## 키워드 변경

정확한일치  대소문자구분

1

변경할 단어  +  → 수정 단어

2

	TF	N-gram
1	안동	71
2	안동시	32
3	예시	30
4	경북	20
5	영탁	18
6	경상북도	12
7	실거	9
8	기부	9
9	music회	8
10	아파트	8
11	계야	8
12	과학	8
13	고향	8
14	주소	8
15	맛집	8
16	출연료	7
17	최고	7
18	새해	7
19	영업시간	7
20	카페	7
21	오전	7
22	행교	6
23	안	6
24	개통	6
25	외	6
26	여행	6
27	길	6
28	전역	6
29	부	6
30	부	6

<TF>

	TF	N-gram
1	경북	안동시 14
2	경상북도	안동시 9
3	계야	music회 8
4	고향	안동 6
5	보	행교 6
6	영탁	고향 6
7	출연료	전역 6
8	기부	계야 5
9	오전	보 4
10	안동	맛집 4
11	아파트	실거 4
12	외	동시 4
13	주소	경북 4
14	기부	영탁 4
15	안동	여행 4
16	동시	기부 4
17	안동	센터 4
18	과학	외 4
19	내년	오전 4
20	센터	자이 4
21	오카	통 4
22	안	동시 4
23	행	사랑 4
24	행교	개통 4
25	생명	과학 3
26	점사	후 3
27	지구	과학 3
28	안동	두 3
29	동시	아파트 3
30	경상북도	안동 3

<N-Gram>

1 좌측 칸에 [변경하고자 하는 단어]를 입력하고

우측 칸에는 [수정된 결과 단어]를 입력합니다.

※ 변경할 단어 여러 개를 하나의 수정단어로 변경하고자 할 때, +(플러스)버튼을 클릭하면 변경할 단어를 추가할 수 있습니다.

※ 단어빈도와 N-gram 분석결과를 참고하면 수정할 유의어, 제거할 불용어, 결합할 단어 찾기에 효과적입니다.

2 탭을 클릭해 TF와 N-Gram 결과를 확인할 수 있습니다.



1



2



## 1 [추천단어] 영역

- N-Gram 기반 확률 모델을 통해 단어쌍을 선별하여 추천합니다.
- 추천단어 리스트에서 변경하고자 하는 추천단어를 클릭하면 입력을 하지 않아도 바로 수정 됩니다.

## 2 [수정내역] 영역

- 수정한 단어 내역을 보여줍니다. 단어를 클릭하면 해당 단어의 변경을 취소할 수 있습니다.
- 단, 변경한 내용 적용 버튼을 누른 후에는 수정 내역을 되돌릴 수 없습니다.

### 개인정보보호

정제정보 확인 복사 다운

정제버전	1	변경단어	개
사용자 사전	0개	정제일시	2023-03-23

분석정보 확인 복사 다운

정제버전	형태소분석기	명사	형용사	동사	외국어	숫자	합계
1	Mecab	77095개	1355개	4381개	1749개	1603개	86183개

데이터 분석 내역 새로고침 복사 다운

구분	세부기능	상태	분석일시	바로가기
단어분석	단어빈도	-		<a href="#">분석하기</a>
	TF-IDF	-		<a href="#">분석하기</a>
	N-gram	-		<a href="#">분석하기</a>
	개체명 인식	-		<a href="#">분석하기</a>
	시계열 분석: 기간별 단어빈도	-		<a href="#">분석하기</a>
	시계열 분석: 기간별 수집량	-		<a href="#">분석하기</a>
매트릭스	매트릭스 단어	-		<a href="#">분석하기</a>
	매트릭스 결과	-		<a href="#">분석하기</a>
	매트릭스 속성 분석	-		<a href="#">분석하기</a>

키워드검색

조회결과 5 건 전체선택 삭제

데이터명	수정날짜	만든날짜	버전
개인정보보호	23-03-23	23-03-23	1
메타버스2	23-03-23	23-03-23	1
메타버스	23-03-22	23-03-17	2
교육부	23-03-22	23-02-22	1
외식	23-03-22	23-02-22	1

- ① 우측 영역은 분석리스트를 조회한 결과를 나타냅니다.
- ② 좌측 영역은 선택한 분석리스트의 정보를 나타냅니다.

☰ 데이터분석
권해진님 ▶ 용량 추가

### 개인정보보호

정제정보 확인 복사 | 다문

정제버전	1	변경단어	개
사용자 사진	0개	정제일시	2023-03-23

분석정보 확인 복사 | 다문

정제버전	형태소분석기	명사	형용사	동사	외국어	숫자	합계
1	Mecab	77095개	1355개	4381개	1749개	1603개	86183개

데이터 분석 내역

- 구분
- 단어분석
- 매트릭스

키워드검색 🔍

조회결과 5 건 전체선택 | 삭제

<input type="checkbox"/>	데이터명	수정날짜	만든날짜	버전
<input type="checkbox"/>	개인정보보호	23-03-23	23-03-23	1
<input type="checkbox"/>	메타버스2	23-03-23	23-03-23	1
<input type="checkbox"/>	메타버스	23-03-22	23-03-17	2
<input type="checkbox"/>	교육부	23-03-22	23-02-22	1
<input type="checkbox"/>	외식	23-03-22	23-02-22	1

키워드검색 **①** 🔍

조회결과 5 건 **②** 전체선택 | 삭제

<input type="checkbox"/>	데이터명	수정날짜	만든날짜	버전
<input type="checkbox"/>	개인정보보호	23-03-23	23-03-23	1
<input type="checkbox"/>	메타버스2	23-03-23	23-03-23	1
<input type="checkbox"/>	메타버스	23-03-22	23-03-17	2
<input type="checkbox"/>	교육부	23-03-22	23-02-22	1
<input type="checkbox"/>	외식	23-03-22	23-02-22	1

**①** 키워드를 검색하여 분석리스트를 조회할 수 있습니다.

**②** 분석리스트를 **[전체선택]** 하거나, 선택한 분석리스트를 **[삭제]** 할 수 있습니다.

✎ 개인정보보호 ①

정제정보 확인

📄 복사 📄 다운

정제버전	1 ▾	② 변경단어	개
사용자 사전	0개	정제일시	2023-03-23

분석정보 확인

📄 복사 📄 다운

정제버전	형태소분석기	명사	형용사	동사 ③	외국어	숫자	합계
1	Mecab	77095개	1355개	4381개	1749개	1603개	86183개

데이터 분석 내역

🔄 새로고침 📄 복사 📄 다운

구분	세부기능	상태	분석일시	바로그기
단어분석	단어빈도	-		분석하기
	TF-IDF	-		분석하기
	N-gram	-		분석하기
	개체명 인식	④		분석하기
	시계열 분석: 기간별 단어빈도	-		분석하기
	시계열 분석: 기간별 수집량	-		분석하기
매트릭스	매트릭스 단어	-		분석하기
	매트릭스 결과	-		분석하기
	매트릭스 속성 분석	-		분석하기

① 데이터명을 나타냅니다. ✎ 클릭하면, 데이터명을 수정할 수 있습니다.

② [정제정보 확인] 에서 해당 데이터를 정제한 조건을 확인할 수 있습니다.

③ [분석정보 확인] 에서 해당 데이터를 형태소분석 결과를 확인할 수 있습니다.

④ [데이터 분석 내역] 에서 해당 데이터로 분석을 진행한 내역을 확인할 수 있습니다.

## 단어빈도(Term Frequency)

1 분석 알고리즘

단어가 문서 내에서 얼마나 자주 발생하는지를 나타내는 척도입니다.  
단어 발생빈도를 중요도 분석을 거쳐 문서의 주제 또는 문서에 대한 태도나 감성을 추론할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

📄 텍스트 다운로드 📄 엑셀 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
외식	335	3.244%	3.244%
포장	301	2.915%	6.16%
맛집	264	2.557%	8.717%
역	245	2.373%	11.09%
중	150	1.452%	12.543%
하	143	1.385%	13.928%
있	135	1.307%	15.236%
집	131	1.268%	16.505%
가족	107	1.036%	17.541%
것	81	0.784%	18.326%

## 데이터 정보

👁️ 원문 📄 정제

외식

수집 날짜: 22.04.06 용량: 615.00 KB

## 시각화 결과

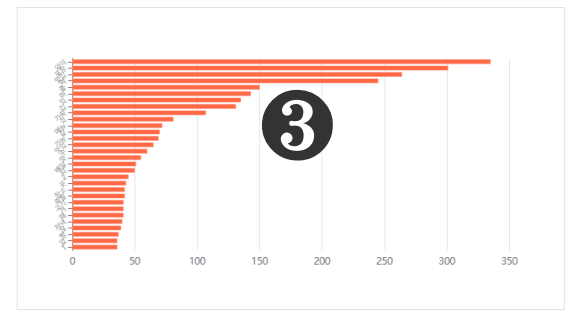
🔍 크게보기 📄 다운로드

가로막대그래프

세로막대그래프

워드클라우드

에고네트워크



시각화 설정

데이터 분석은 데이터 분석 콘솔에서 진행됩니다.

데이터 분석 콘솔은 크게 3개의 영역으로 구성되어 있습니다.

① 영역은 분석에 적용된 **[데이터 정보]**를 나타냅니다.

② 영역은 **[데이터 분석 결과]**를 나타냅니다.

③ 영역은 데이터 분석의 **[시각화 결과]**를 나타냅니다.

## 단어빈도(Term Frequency)

📌 분석 알고리즘

단어가 문서 내에서 얼마나 자주 발생하는지를 나타내는 척도입니다.  
단어 발생빈도를 중요도 분석을 거쳐 문서의 주제 또는 문서에 대한 태도나 감성을 추론할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

단어	빈도(회)	백분	
외식	335	3.2	
포장	301	2.9	
맛집	264	2.5	
역	245	2.3	
출	150	1.4	
하	143	1.3	
있	135	1.307%	15.236%
집	131	1.268%	16.505%
가족	107	1.036%	17.541%
것	81		

## 데이터 정보

👁 원문

👁 정제

외식

수집 날짜 : 22.04.06 용량 : 615.00 KB

## 시각화 결과

🔍 크게보기

📄 다운로드

## 데이터 정보

1

👁 원문

👁 정제

외식

수집 날짜 : 22.04.06 용량 : 615.00 KB

## 데이터 상세 정보

✕

2

데이터명	외식
용량	615.00 KB
수집날짜	22.04.06
형태소분석기	mecab
분석품사	NN,NNG,NNP,NNB,VA,VV,SL,SN
작업날짜	23.03.22

닫기

1 [데이터 정보] 에서 분석에 적용된 데이터 정보를 확인할 수 있습니다.

2 데이터 정보 화살표를 클릭하면, [데이터 상세 정보] 가 나타납니다.

단어빈도(Term Frequency)

단어	빈도(회)	백분
외식	335	3.2%
포장	301	2.9%
맛길	264	2.5%
먹	245	2.3%
총	150	1.4%
하	143	1.3%
있	135	1.307%
김	131	1.268%

데이터 정보

외식  
수집 날짜: 22.04.06 용량: 615.00 KB

원문 정제

원문 / 정제 데이터 미리보기

데이터명	생성일자	용량
untitled		
외식	2023-03-22	615.00 KB

원문데이터 정제데이터

19년 9월 가계부 (저축률 73.5%) 원래 요리를 좋아하니 별로 어려울 줄다. 건강하고 내 입맛대로 먹고 양도 남기는.. naver blog ht

10월 양산천태산 테마 산행안내 덕천동외식1번가건편 세븐일레븐편입니다. 산행코스 및 산소개 1. 산행코스 q=%22%EC%99%B8%EC%8

★ 용인시 주택 경매 ♥ 경기 용인시 수지구 신봉동... 단지, 임야, 농경수목, 바닥포장, 화단, 옹벽 등은 ㅇ 토지의 부합 q=%22%EC%99%B8%EC%8

3 마시 잔수 있선 포장 약규정 것 것 인생 안약도모유익 주 연고

텍스트 (.txt) 엑셀 (.xlsx)

③ 원문 정제 버튼을 클릭하면, 분석데이터의 [원문/정제 데이터 미리보기] 가 팝업으로 나타납니다.

④ 텍스트 (.txt) 엑셀 (.xlsx) 버튼을 클릭하면, 미리보기 창의 데이터를 해당 파일 형식으로 다운받을 수 있습니다.

※ 분석에 적용된 [원문 데이터]의 텍스트(.txt) 파일, 엑셀(.xlsx) 파일 다운로드

※ 분석에 적용된 [정제 데이터]의 텍스트(.txt) 파일, 엑셀(.xlsx) 파일 다운로드

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

## 단어빈도(Term Frequency)

**1** 분석 알고리즘

단어 빈도는 단어가 문서 내에서 얼마나 자주 발생하는지를 나타내는 척도입니다. 빈도 분석은 특정 문서들 내에서 자주 사용되는 단어를 추출하여 언급되는 빈도수에 따라 중요도를 분석하는 방법으로 발생빈도를 파악하여 문서의 주제 또는 문서에 대한 태도나 감성을 추론할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

**2** 텍스트 다운로드 엑셀 다운로드

<b>3</b>	단어	빈도(건)	백분율(%)	누적백분율(%)
	외식	335	3.244%	3.244%
	포장	301	2.915%	6.16%
	맛집	264	2.557%	8.717%
	먹	245	2.373%	11.09%
	중	150	1.452%	12.543%
	하	143	1.385%	13.928%
	있	135	1.307%	15.236%
	집	131	1.268%	16.505%
	가족	107	1.036%	17.541%
	것	81	0.784%	18.326%

**1** [분석 알고리즘] 버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** 단어빈도 분석결과로 추출된 단어와 데이터 내 해당 단어의 빈도(건), 백분율(%), 누적백분율(%)을 나타냅니다.

- 백분율(%)은 전체 언급량을 100으로 했을 때 언급된 양을 의미합니다.
- 빈도(건) 수가 높다는 것은 정제데이터 내에 해당 단어가 등장하는 빈도가 높다는 것을 의미합니다.



데이터 정보 
원문 
정제

untitled

수집 날짜: 2022-04-06   용량: 615 KB

1 
2

시각화 결과

가로막대그래프
세로막대그래프
워드클라우드
3

시각화 설정 **4**

▶ 색상선택

빈도 ● 백분율 ●  데이터 표시

▶ 키워드 선정

✓ 선택해제
🏆 상위 50개

빈도	키워드	내림차순	오름차순
1~50		51~100	
<input checked="" type="checkbox"/> 외식	335	<input type="checkbox"/> 3	23
<input checked="" type="checkbox"/> 포장	301	<input type="checkbox"/> 층	23
<input checked="" type="checkbox"/> 맛집	264	<input type="checkbox"/> 주차	23

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 단어빈도 분석은 4종의 시각화 결과를 제공합니다.

- 가로막대 그래프, 세로막대 그래프, 워드클라우드, 에고네트워크

데이터 정보 ↗

원문 | 경제

untitled

수집 날짜: 2022-04-06 용량: 615 KB

시각화 결과

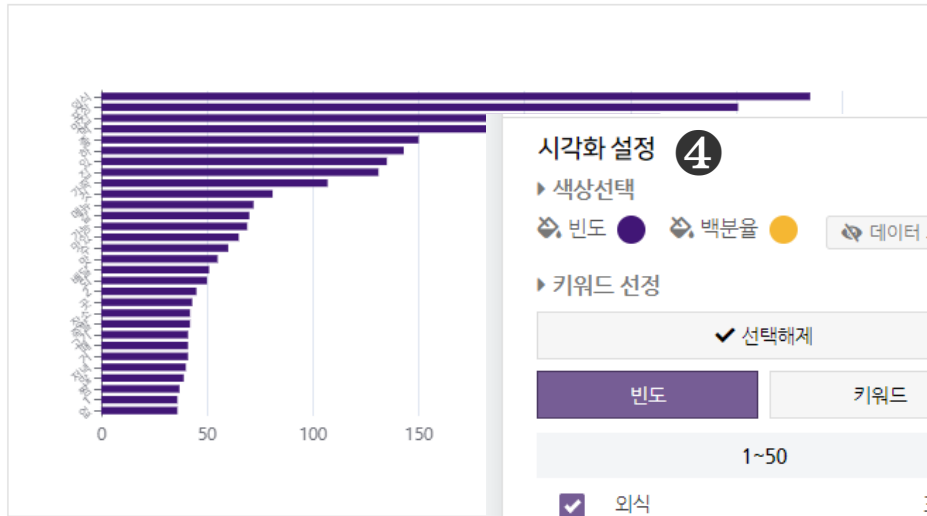
크게보기 | 다운로드

가로막대그래프

세로막대그래프

워드클라우드

에고네트워크



## 시각화 설정 4

▶ 색상선택

빈도 (purple) | 백분율 (yellow) | 데이터 표시

▶ 키워드 선정

선택해제

상위 50개

빈도

키워드

내림차순

오름차순

1~50

51~100

키워드	빈도	순위
<input checked="" type="checkbox"/> 외식	335	<input type="checkbox"/> 3
<input checked="" type="checkbox"/> 포장	301	<input type="checkbox"/> 층
<input checked="" type="checkbox"/> 맛집	264	<input type="checkbox"/> 주차

시각화 설정

④ [시각화 설정] 을 통해시각화의 색상과 키워드를 변경할 수 있습니다.

- 빈도 : 막대 색상, 백분율 : 꺾은선 색상

- 키워드 : 빈도 또는 키워드(내림차순/오름차순) 기준으로 정렬하여 키워드를 선택할 수 있습니다.

※ 막대 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다

단어빈도

**N-gram**

TF-IDF

개체명 인식

시계열

## N-gram(N-gram)

**1** [분석 알고리즘](#)

N-gram은 문자열에서 N개의 연속된 요소를 추출하는 통계 기반 언어 분석모델로, 문장 내 연속하여 동시 출현하는 단어와 그 빈도를 계산합니다. N-gram은 문장을 몇 개의 단어 개수에 따라 나눌지에 따라 종류가 결정되는데, 텍스트용은 N=2인 Bigram 모델을 제공합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

**2** [텍스트 다운로드](#) [엑셀 다운로드](#)

<b>3</b>	단어1	단어2	빈도(건)
	가족	외식	85
	수	있	44
	것	같	43
	포장	가능	38
	외식	하	30
	메뉴	포장	20
	외식	총	19
	외식	장소	19
	총	것	17
	먹	수	17

**1** [\[분석 알고리즘\]](#)을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [\[텍스트 다운로드\]](#), [\[엑셀 다운로드\]](#) 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** N-gram 분석결과로 N개의 연속된 단어(단어쌍)와 그 빈도(건) 을 산출합니다.

- 단어1, 단어2(단어쌍)의 빈도가 높다는 것은 두 단어가 동시에 등장하는 빈도가 높다는 것을 의미합니다.

※ 텍스트용은 N=2인 Bi-gram 값을 제공하고 있습니다.

데이터 정보

원문 | 정제

외식

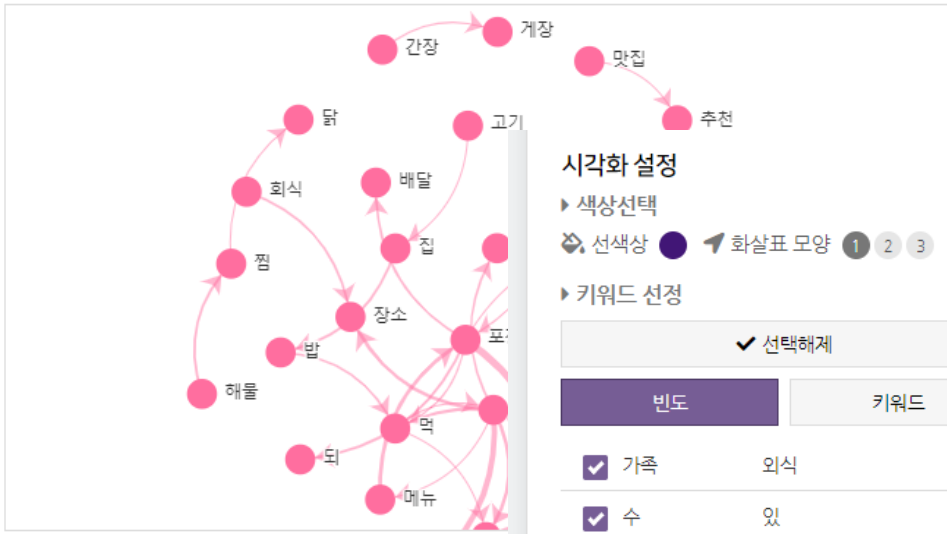
수집 날짜 : 22.04.06 용량 : 615.00 KB

시각화 결과

크게보기

다운로드

N-gram



시각화 설정

색상선택

선택상 | 화살표 모양 1 2 3

키워드 선정

선택해제

상위 50개

빈도

키워드

내림차순

오름차순

<input checked="" type="checkbox"/>	가족	외식	85
<input checked="" type="checkbox"/>	수	있	44
<input checked="" type="checkbox"/>	것	같	43
<input checked="" type="checkbox"/>	포장	가능	38

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ N-gram 분석은 1종의 시각화 결과를 제공합니다.

데이터 정보

원문 | 정제

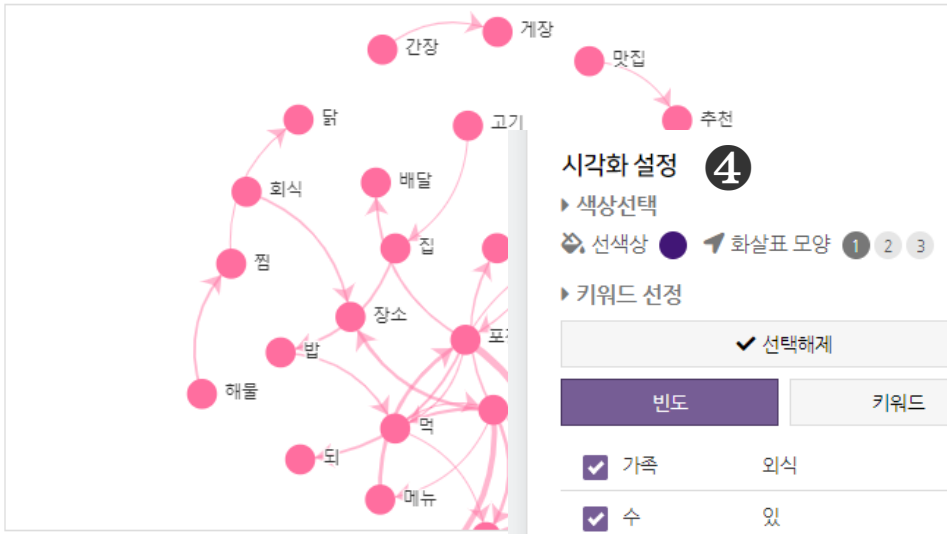
외식

수집 날짜 : 22.04.06 용량 : 615.00 KB

시각화 결과

크게보기 | 다운로드

N-gram



### 시각화 설정 4

▶ 색상선택

선택상 | 화살표 모양 1 2 3

▶ 키워드 선정

선택해제

상위 50개

빈도

키워드

내림차순

오름차순

<input checked="" type="checkbox"/>	가족	외식	85
<input checked="" type="checkbox"/>	수	있	44
<input checked="" type="checkbox"/>	것	같	43
<input checked="" type="checkbox"/>	포장	가능	38

시각화 설정

④ [시각화 설정]을 클릭하면 시각화의 선 색상, 화살표 모양과 키워드를 변경할 수 있습니다.

- 빈도 또는 키워드(내림차순/오름차순) 기준으로 정렬하여 키워드를 선택할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

※ 단어 위에 마우스를 올리면 연결된 단어 구조를 확인할 수 있습니다

단어빈도

N-gram

**TF-IDF**

개체명 인식

시계열

## TF-IDF(Term Frequency-Inverse Document Frequency)

**1** ⓘ 분석 알고리즘

TF-IDF는 단어의 빈도와 역문서 빈도를 사용하여 문서 내의 각 단어에 가중치를 부여하는 것으로, 문서 내 단어가 얼마나 중요한지를 평가합니다. TF-IDF값이 높은 단어는 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높고, TF-IDF값이 낮은 단어는 문서 내에서 흔히 사용되고 있을 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

**2**

📄 텍스트 다운로드 📄 엑셀 다운로드

**3**

단어	TF-IDF	DF	IDF
먹	218.36591026884076	162	0.8912894296687378
맛집	189.07563206300642	193	0.7161955759962364
좋	178.70910331786135	120	1.1913940221190757
집	173.5652293314114	105	1.3249254147435985
하	169.18261636052645	121	1.1830952193043809
있	167.76278772839456	114	1.2426873165066263
가족	150.24670214455807	97	1.404174786397739
메뉴	139.38008378879317	57	1.9358344970665717
것	137.8797913166904	72	1.7022196458850665
가능	128.89282115021894	61	1.8680119007278106

**1** [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** TF-IDF 분석결과로 단어의 TF-IDF, DF, IDF 값을 산출합니다.

- TF-IDF : TF와 IDF를 곱한 값으로, 문서 내 단어가 얼마나 중요한지를 평가하는 지표
- TF(Term Frequency) : 전체 문서 내에서 단어 w의 등장 횟수
- DF(Document Frequency) : 단어 w가 등장한 문서의 수
- IDF(Inverse Document Frequency) : DF의 역수, (전체 문서 수)/(해당 단어가 등장한 문서 수)

데이터 정보 ↗
원문    경제

---

untitled  
수집 날짜 : 22.04.06    용량 : 615 KB

---

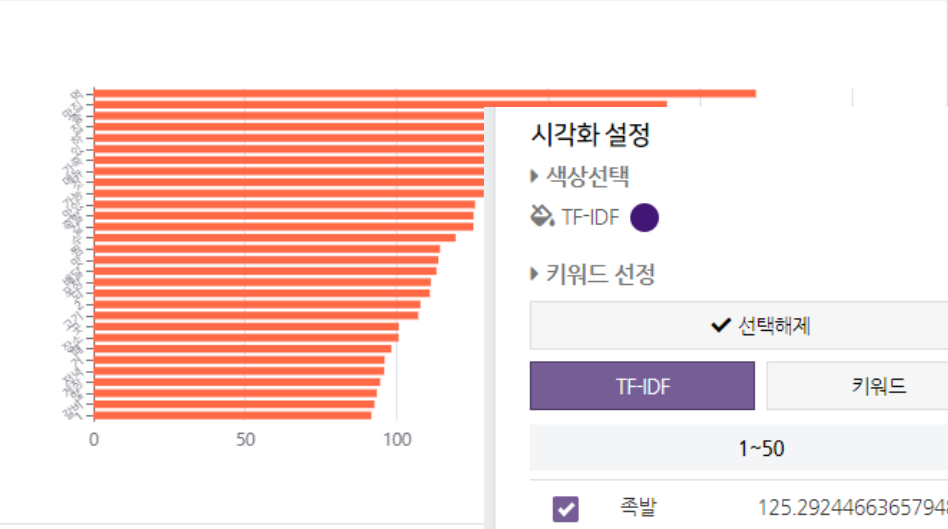
시각화 결과
크게보기    다운로드

---

가로막대그래프
세로막대그래프
워드클라우드
에고네트워크

3

---



**시각화 설정**

▶ 색상선택

TF-IDF ●

▶ 키워드 선정

선택해제

5 상위 50개

TF-IDF

키워드

내림차순

오름차순

1~50

51~100

<input checked="" type="checkbox"/>	족발	125.29244663657948	<input type="checkbox"/>	방문	69.00024850082302
<input checked="" type="checkbox"/>	갈	125.24617160122872	<input type="checkbox"/>	엄마	68.68754683977716
<input checked="" type="checkbox"/>	수	119.39410310021086	<input type="checkbox"/>	주차	68.612530300984

---

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ TF-IDF 분석은 4종의 시각화 결과를 제공합니다.

- 가로막대 그래프, 세로막대 그래프, 워드클라우드, 에고네트워크

데이터 정보

원문    경제

untitled

수집 날짜 : 22.04.06    용량 : 615 KB

시각화 결과

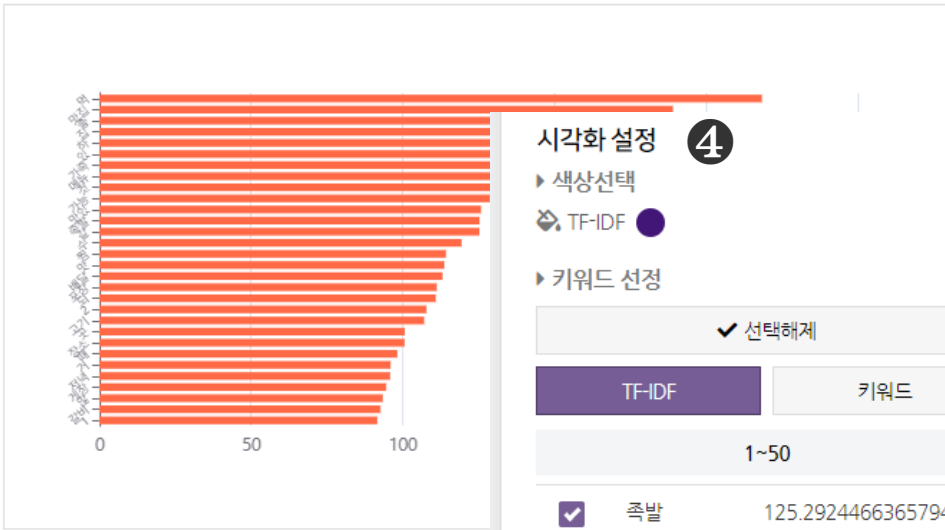
크게보기    다운로드

가로막대그래프

세로막대그래프

워드클라우드

에고네트워크



## 시각화 설정 ④

▶ 색상선택

TF-IDF

▶ 키워드 선정

선택해제

상위 50개

TF-IDF

키워드

내림차순

오름차순

1~50

51~100

<input checked="" type="checkbox"/>	족발	125.29244663657948
<input checked="" type="checkbox"/>	갈	125.24617160122872
<input checked="" type="checkbox"/>	수	119.39410310021086

<input type="checkbox"/>	방문	69.00024850082302
<input type="checkbox"/>	엄마	68.68754683977716
<input type="checkbox"/>	주차	68.612530300984

시각화 설정

## ④ [시각화 설정]

- 시각화의 색상과 키워드를 변경할 수 있습니다.
- TF-IDF : 막대 색상
- TF-IDF 또는 키워드(내림차순/오름차순) 기준으로 정렬하여 키워드를 선택할 수 있습니다.

※ 막대 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다



단어빈도

N-gram

TF-IDF

개체명 인식

시계열

## 개체명 인식(Named Entity Recognition)

1 [분석 알고리즘](#)

미리 정의해 놓은 개체명을 문서에서 인식하여 추출, 분류하는 기법입니다.  
 텍스트에서는 사람에서부터 용어에 이르기까지 14개의 개체명을 인식하여 제공합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 [텍스트 다운로드](#) [엑셀 다운로드](#)

사람	학문	대상물	기관	지역	문명	날짜	시간	숫자	사건/사고	동물	식물	금속	용어
				단어									빈도(건)
				청주									10
				일산									9
				용인									9
				송도									8
				부산									7
				마곡									6
				서울									6
				교동									5

1 **[분석 알고리즘]**을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 **[텍스트 다운로드]**, **[엑셀 다운로드]** 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

## 개체명 인식(Named Entity Recognition)

[분석 알고리즘](#)

미리 정의해 놓은 개체명을 문서에서 인식하여 추출, 분류하는 기법입니다.  
 텍스트에서는 사람에서부터 용어에 이르기까지 14개의 개체명을 인식하여 제공합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

사람	학문	대상물	기관	지역	문명	날짜	시간	<b>3</b> 숫자	사건/사고	동물	식물	금속	용어
----	----	-----	----	----	----	----	----	-------------	-------	----	----	----	----

단어	빈도(건)
청주	10
일산	9
용인	9
송도	8
부산	7
마곡	6
서울	6
교동	5

### ③ 개체명 인식 분석결과로 14개 개체명 범주에 따라 분류된 단어와 그 빈도(건)를 산출합니다.

※ 개체명(Named Entity) 범주

- 사람, 학문, 대상물, 기관, 지역, 문명, 날짜, 시간, 숫자, 사건/사고, 동물 식물, 금속, 용어

※ 텍스트롬은 창원대학교에서 개발한 개체명 인식기로 분석하고 있습니다.

※ 산출되는 단어는 개체명 인식기에 탑재된 형태소 분석기로 출력된 것으로, 정제/편집한 데이터와는 차이가 있습니다.

※ 해당 문서에서 14개 개체명에 속하는 키워드가 없다면 개체명이 나타나지 않습니다.

데이터 정보

원문 | 정제

외식

수집 날짜 : 22.04.06 | 용량 : 615.00 KB

시각화 결과

크게보기

다운로드

트리맵



시각화 설정

▶ 색상선택

- 사람
- 학문
- 대상물
- 기관
- 지역
- 문명
- 날짜
- 시간
- 숫자
- 사건/사고
- 동물
- 식물
- 금속
- 용어

시각화 설정

- ① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.
- ② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.
- ③ 개체명 인식 분석은 1종의 시각화 결과(트리맵)를 제공합니다.
- ④ [시각화 설정]을 통해, 시각화의 색상을 변경할 수 있습니다.
  - ※ 맵 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다
  - ※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

시계열 분석: 기간별 수집량

시계열 분석: 기간별 단어빈도

### 시계열 분석: 기간별 수집량

1 분석 알고리즘

수집단위에서 설정한 기간 및 수집 채널별로 데이터 수집량을 분석합니다. 시계열분석은 수집하기에서 수집단위를 사용하여 수집한 데이터만 분석이 가능합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

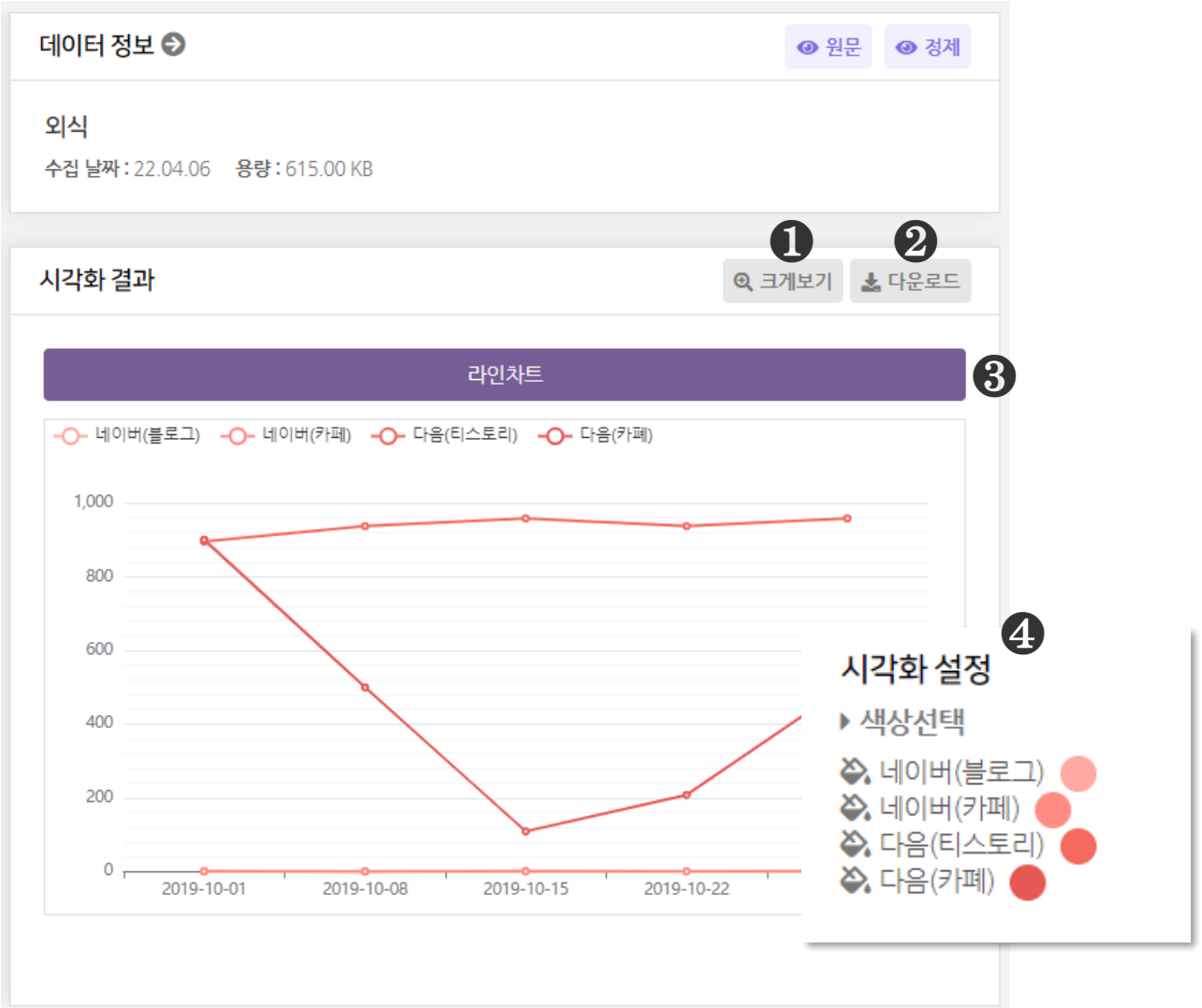
3 키워드	날짜	네이버(블로그)	네이버(카페)	다음(블로그)	다음(카페)
"외식 포장"	2019-10-01 ~ 2019-10-01	2	0	895	900
"외식 포장"	2019-10-08 ~ 2019-10-08	0	0	937	500
"외식 포장"	2019-10-15 ~ 2019-10-15	2	0	958	109
"외식 포장"	2019-10-22 ~ 2019-10-22	1	0	937	208
"외식 포장"	2019-10-29 ~ 2019-10-29	2	0	958	505

1 **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 **[텍스트 다운로드], [엑셀 다운로드]** 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

3 시계열 기간별 수집량 분석결과로 기간별로 수집한 각 채널의 수집량(문서량)을 산출합니다.

※ 데이터 수집 시, 수집단위를 이용하여 수집을 한 경우에만 시계열 분석이 가능합니다.



① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 시계열 기간별 수집량 분석은 1종의 시각화 결과(라인차트)를 제공합니다.

④ [시각화 설정]을 통해, 시각화의 색상을 변경할 수 있습니다.

※ 표식 위에 마우스를 올리면 데이터 값(문서량)을 확인할 수 있습니다.

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

시계열 분석: 기간별 수집량

시계열 분석: 기간별 단어빈도

## 시계열 분석: 기간별 단어빈도

1 ⓘ 분석 알고리즘

수집단위에서 설정한 기간별로, 선택한 단어의 출현 빈도를 분석합니다. 시계열분석은 수집하기에서 수집단위를 사용하여 수집한 데이터만 분석이 가능합니다.

2

▶ 분석단어 선택

바로 선택하기

바로 선택하기 팝업을 통해 분석단어를 선택해보세요.

3

적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드

분석단어선택 (시계열분석)

단어	데이터명	생성일자	용량
	교육부	2023-03-22	830.13 KB

선택단어수  확인 선택단어누적비율  %

상위 200개 까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

<input type="checkbox"/>	단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/>	교육부	1605	5.54480757272162	5.54480757272162
<input type="checkbox"/>	교육	529	1.8275409382989014	7.372348511020522
<input type="checkbox"/>	학교	341	1.178055689905341	8.550404200925863
<input type="checkbox"/>	2	249	0.8602224832446624	9.410626684170525
<input type="checkbox"/>	2021	202	0.6978511711462724	10.108477855316798

다운로드

1 [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 [바로 선택하기] 버튼을 클릭하여 분석할 단어 최대 10개 선택합니다.

3 [적용] 버튼을 눌러, 시계열 분석을 실행해주세요.

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

시계열 분석: 기간별 수집량

시계열 분석: 기간별 단어빈도

시계열 분석: 기간별 단어빈도

[분석 알고리즘](#)

수집단위에서 설정한 기간별로, 선택한 단어의 출현 빈도를 분석합니다. 시계열분석은 수집하기에서 수집단위를 사용하여 수집한 데이터만 분석이 가능합니다.

▶ 분석단어 선택

[바로 선택하기](#)

바로 선택하기 팝업을 통해 분석단어를 선택해보세요.

✓ 적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

④

[텍스트 다운로드](#)

[엑셀 다운로드](#)

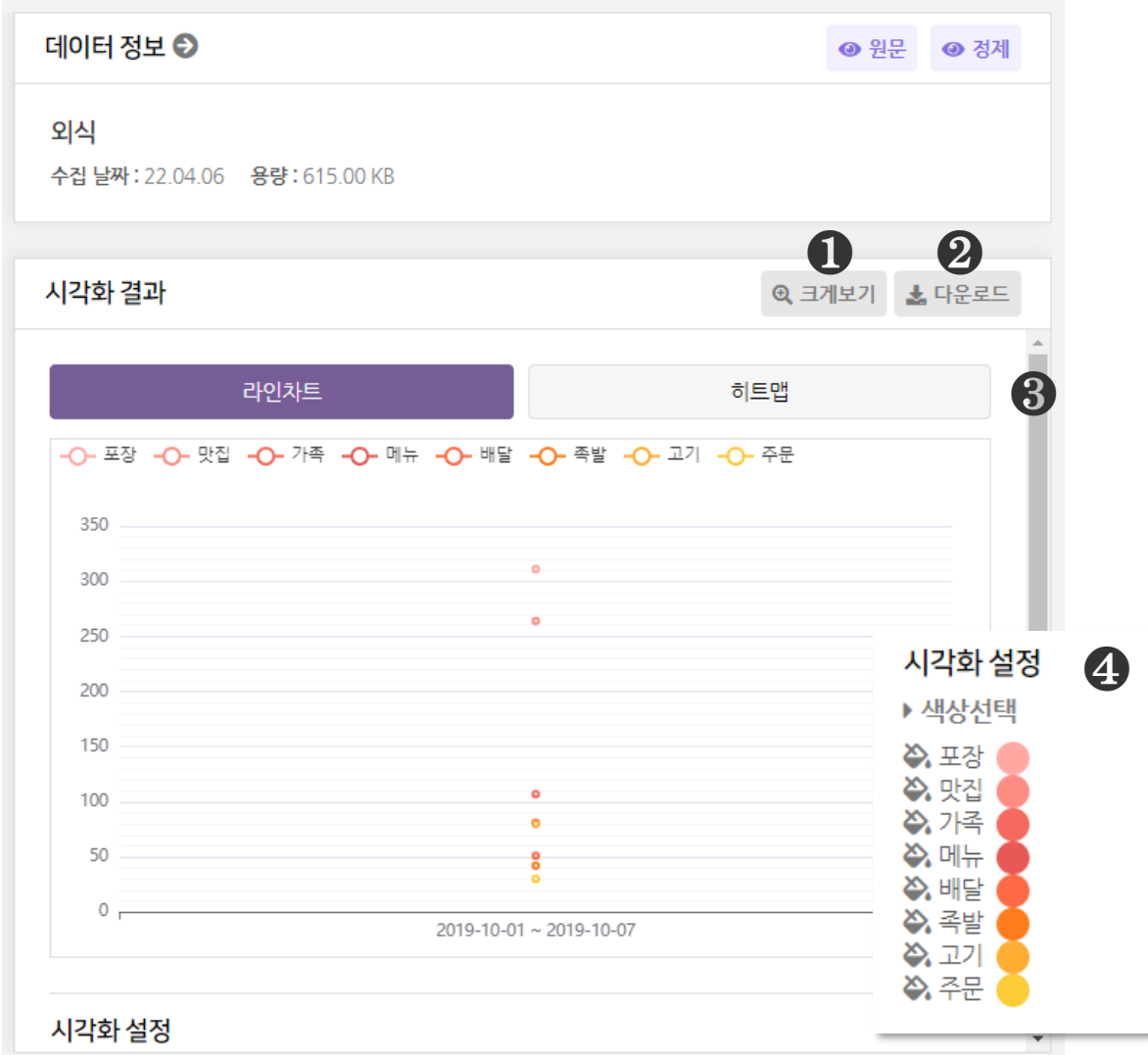
⑤

단어	날짜	빈도
포장	2019-10-01 ~ 2019-10-07	311
맛집	2019-10-01 ~ 2019-10-07	264
가족	2019-10-01 ~ 2019-10-07	107
메뉴	2019-10-01 ~ 2019-10-07	81
배달	2019-10-01 ~ 2019-10-07	51
족발	2019-10-01 ~ 2019-10-07	42
고기	2019-10-01 ~ 2019-10-07	80

④ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

⑤ 시계열 기간별 단어빈도 분석결과로 선택한 단어의 기간별 빈도를 산출합니다.

※ 데이터 수집 시, 수집단위를 이용하여 수집을 한 경우에만 시계열 분석이 가능합니다.



① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 시계열 기간별 단어빈도 분석은 2종의 시각화 결과를 제공합니다.

- 라인차트, 히트맵

④ [시각화 설정] 을 통해, 시각화의 색상을 변경할 수 있습니다.

※ 점(표식) 위에 마우스를 올리면 데이터 값(단어빈도)을 확인할 수 있습니다.



1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 단어(Matrix Word)

① [분석 알고리즘](#)

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다.  
최대 500개 단어를 선택할 있습니다. 200개 이상 선택 시, 파일로 업로드 해주세요.

[예시파일](#)

단어선택

②

단어 선택하기

파일 업로드

③

적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하

번호

#### 분석단어선택 (매트릭스 1-Mode 단어 선택)

데이터명	생성날짜	용량
교육부	2023-03-17	830.1 KB

선택단어수  확인    선택단어누적비율  %    [다운로드](#)    [적용](#)

상위 200개 까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

<input type="checkbox"/>	단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/>	교육부	1546	5.448458149779736	5.448458149779736
<input type="checkbox"/>	교육	502	1.7691629955947137	7.21762114537445
<input type="checkbox"/>	학교	349	1.229955947136564	8.447577092511013
<input type="checkbox"/>	2	246	0.8669603524229075	9.31453744493392
<input type="checkbox"/>	있	210	0.7400881057268722	10.054625550660793
<input type="checkbox"/>	경관	187	0.6590308370044053	10.713656387665198

① [분석 알고리즘]을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② [매트릭스 분석을 진행할 단어를 선택] : 단어 선택하기

- [단어 선택하기] 를 클릭하면, 단어를 선택할 수 있는 새 창이 나타납니다.

③ 단어 선택을 완료한 후, [적용] 버튼을 눌러주세요.

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 단어(Matrix Word)

① [분석 알고리즘](#)

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다.  
최대 500개 단어를 선택할 있습니다. 200개 이상 선택 시, 파일로 업로드 해주세요.

[예시파일](#)

▶ 단어선택

②

[단어 선택하기](#)

[파일 업로드](#)

행단어 파일을 업로드해 보세요.



③

[적용](#)

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

번호

단어

빈도(건)

분석 결과가 존재하지 않습니다.

① **[분석 알고리즘]**을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다

② **[매트릭스 분석을 진행할 단어를 선택] : 파일 업로드**

- [파일 업로드] 를 클릭하면, 오른쪽에서 파일 업로드 창이 생성됩니다.

- 엑셀(xlsx), 메모장(txt)을 활용해 단어 목록을 작성해 업로드 해주세요.

※ 파일 업로드 시, 데이터에 등장하는 단어를 입력해주세요.

③ 단어 선택을 완료한 후, **[적용]** 버튼을 눌러주세요.

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

매트릭스 단어(Matrix Word)

[분석 알고리즘](#)

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다.  
최대 500개 단어를 선택할 있습니다. 200개 이상 선택 시, 파일로 업로드 해주세요.

[예시파일](#)

▶ 단어선택

[단어 선택하기](#)

[파일 업로드](#)

행단어 파일을 업로드해 보세요.

[📎](#)

✓ 적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

④

[텍스트 다운로드](#) [엑셀 다운로드](#)

⑤

번호	단어	빈도(건)
1	외식	335
2	포장	301
3	맛집	264
4	가족	107
5	메뉴	72
6	맛있	65

④ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

⑤ 매트릭스 단어 결과로 매트릭스 분석에 선택한 단어의 빈도가 나타납니다.

데이터 정보 → 원문 경제

untitled  
수집 날짜: 22.04.06 용량: 615 KB

시각화 결과 크게보기 다운로드

가로막대그래프 세로막대그래프 워드클라우드

시각화 설정

- 색상선택
- 빈도
- 키워드 선정

선택해제	상위 50개
빈도	내림차순
키워드	오름차순
1~50	51~100
<input checked="" type="checkbox"/> 외식 335	
<input checked="" type="checkbox"/> 포장 301	
<input checked="" type="checkbox"/> 맛집 264	

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 매트릭스 단어는 3종의 시각화 결과를 제공합니다.

- 가로막대 그래프, 세로막대 그래프, 워드클라우드

④ [시각화 설정] 을 통해, 시각화의 색상과 키워드를 변경할 수 있습니다.

- 빈도 : 막대 색상

- 빈도 또는 키워드(내림차순/오름차순) 기준으로 정렬하여 키워드를 선택할 수 있습니다.

※ 막대 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 결과(Matrix Result)

1 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스

엣지리스트

유클리디언 계수

2

코사인 계수

자카드 계수

상관 계수

전체 단어는 다운로드하여 확인할 수 있습니다.

3 텍스트 다운로드 엑셀 다운로드

	외식	포장	맛집	가족	메뉴	맛있	배달	족발	주문	갈비	회식	주말	계장	피자	주차	할인	떡볶이	모임	해물
외식	0	243	240	118	66	57	49	38	22	18	27	23	18	15	20	18	17	17	19
포장	243	0	197	96	57	51	46	41	26	29	27	16	24	15	19	27	8	18	14
맛집	240	197	0	103	67	63	38	60	20	36	27	24	32	28	28	20	9	25	15
가족	118	96	103	0	22	22	0	16	6	7	15	0	4	4	0	12	2	0	10

1 [분석 알고리즘]을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 매트릭스 결과(Matrix Result)는 6종의 분석 결과를 매트릭스 형태로 제공합니다.

- 매트릭스, 엣지리스트, 유클리디언 계수, 코사인 계수, 자카드 계수, 상관계수

※ 분석 명칭을 클릭하면, 해당 분석결과를 하단에 나타냅니다.

3 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

## 매트릭스 결과(Matrix Result)

### ④ 매트릭스 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스    옛지리스트    유클리디언 계수    코사인 계수    자카드 계수    상관 계수

전체 단어는 다운로드하여 확인할 수 있습니다.

텍스트 다운로드   엑셀 다운로드

	외식	포장	맛집	가족
외식	0	243	240	118
포장	243	0	197	96
맛집	240	197	0	103
가족	118	96	103	0

## 매트릭스 결과(Matrix Result)

### ⑤ 옛지리스트 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스    **옛지리스트**    유클리디언 계수    코사인 계수    자카드 계수    상관 계수

전체 단어는 다운로드하여 확인할 수 있습니다.

텍스트 다운로드   엑셀 다운로드

Word1
외식
외식
외식
외식
외식
외식
외식
외식
외식

## 매트릭스 결과(Matrix Result)

### ⑥ 유클리디언 계수 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스    옛지리스트    **유클리디언 계수**    코사인 계수    자카드 계수    상관 계수

전체 단어는 다운로드하여 확인할 수 있습니다.

텍스트 다운로드   엑셀 다운로드

외식	포장	맛집	가족	메뉴	맛있	배달	족발	주문	갈비	회식	주말
0.0	0.8571	0.9028	0.6837	0.7574	0.6666	0.7226	0.8599	0.4226	0.5	0.4226	0.2928
0.8571428571428572	0	0.8829	0.8174	0.7574	0.6464	0.7574	0.8259	0.5527	0.7817	0.6837	0
0.9028714137642736	0.8829	0	0.8686	0.8143	0.8377	0.811	0.8285	0.6984	0.6984	0.5917	0.8333
0.683772233983162	0.8174	0.8686	0	0.2928	0.5527	0	0.7705	0	0.4226	0.2928	0.2928
0.757464374963667	0.7574	0.8143	0.2928	0	0.2928	0.5917	0.5	0.5917	0	0	0.2928
0.6666666666666667	0.6464	0.8377	0.5527	0.2928	0	0	0.811	0	0.4226	0	0
0.7226499018873854	0.7574	0.811	0	0.5917	0	0	0.8	0.6984	0	0	0
0.8599719915971991	0.8259	0.8285	0.7705	0.5	0.811	0.8	0	0	0	0	0

④ [매트릭스] 는 전체 문서 내에서 매트릭스 선택단어의 출현빈도가 높은 순서대로 단어와 빈도를 표시합니다.

⑤ [옛지리스트] 는 전체 문서 내에서 매트릭스 선택단어의 연결 쌍, 즉 단어와 단어를 짝지어 목록으로 나타냅니다. (Weight : 빈도)

⑥ [유클리디언 계수] 는 유클리디언 거리(Euclidean Distance)를 바탕으로 유사도를 나타낸 지표입니다.

※ 유사도 값은 0과 1 사이의 값을 갖고, 1에 가까울수록 두 단어의 거리가 가깝다는 것을 의미

## 매트릭스 결과(Matrix Result)

### ⑦ 코사인 계수 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스	옛지리스트	유클리디언 계수	<b>코사인 계수</b>	자카드 계수	상관 계수
------	-------	----------	---------------	--------	-------

전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

외식	포장
0.0	0.5447
0.5447474541731129	0
0.3962977363363993	0.3363
0.43308618881685385	0.341
0.23417016625941414	0.2216
0.2761495323085321	0.2151
0.21210563101261748	0.2082
0.08172699125821396	0.068

## 매트릭스 결과(Matrix Result)

### ⑧ 자카드 계수 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스	옛지리스트	유클리디언 계수	코사인 계수	<b>자카드 계수</b>	상관 계수
------	-------	----------	--------	---------------	-------

전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

외식	포장
0.0	0.316
0.3160377358490566	0
0.2704507512520868	0.2407
0.2149321266968326	0.1813
0.11547911547911548	0.1179
0.12	0.101
0.09067357512953368	0.0965
0.04509283819628647	0.0408

## 매트릭스 결과(Matrix Result)

### ⑨ 상관 계수 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. 최대 500개 단어를 선택할 있습니다.  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스	옛지리스트	유클리디언 계수	코사인 계수	자카드 계수	<b>상관 계수</b>
------	-------	----------	--------	--------	--------------

전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

외식	포장	맛집	가족	메뉴	맛있	배달	족발	주문	갈비	회식	주
0.0	0.1118	0.1036	0.289	0.0537	0.0236	0.0747	0.0239	0.0611	0.1021	0.0293	0.03
0.11182767283293382	0	0.0217	0.1241	0.0188	0.015	0.0751	0.0731	0.0451	0.0682	0.0631	0.01
0.1036369699050652	0.0217	0	0.191	0.1176	0.1415	0.0291	0.1834	0.0005	0.1252	0.0748	0.01
0.28904608386928504	0.1241	0.191	0	0.0358	0.0523	0.059	0.0437	0.0355	0.0145	0.122	0.01
0.05375793545179941	0.0188	0.1176	0.0358	0	0.0103	0.0214	0.0519	0.0606	0.0336	0.0996	0.01
0.023698909642772888	0.015	0.1415	0.0523	0.0103	0	0.0496	0.1703	0.0386	0.0627	0.041	0.01
0.07479596769367639	0.0751	0.0291	0.059	0.0214	0.0496	0	0.2158	0.2079	0.0615	0.0408	0.01
0.023900660730759124	0.0731	0.1834	0.0437	0.0519	0.1703	0.2158	0	0.0128	0.039	0.02	0.01

⑦ [코사인 계수] 는 코사인 유사도로, 단어를 벡터로 표현하여 벡터간 이루는 내각의 크기로 유사도를 측정하는 것입니다.

※ 유사도 값은 0과 1 사이의 값을 갖고, 1에 가까울수록 유사도가 높습니다.

⑧ [자카드 계수] 는 자카드 유사도로, 두 집합 사이의 겹치는 정도를 나타낸 것입니다.

※ 유사도 값은 0과 1 사이의 값을 갖고, 1에 가까울수록 유사도가 높습니다.

⑨ [상관 계수] 는 두 변수 중에서 한 변수의 변화가 다른 변수의 변화에 따라 어떻게 변화하는지 보여주는 지표입니다.

※ 상관계수의 절대값이 높을수록 두 변수 간의 관계가 높다고 할 수 있습니다.

데이터 정보

원문    정제

외식

수집 날짜 : 22.04.06    용량 : 615 KB

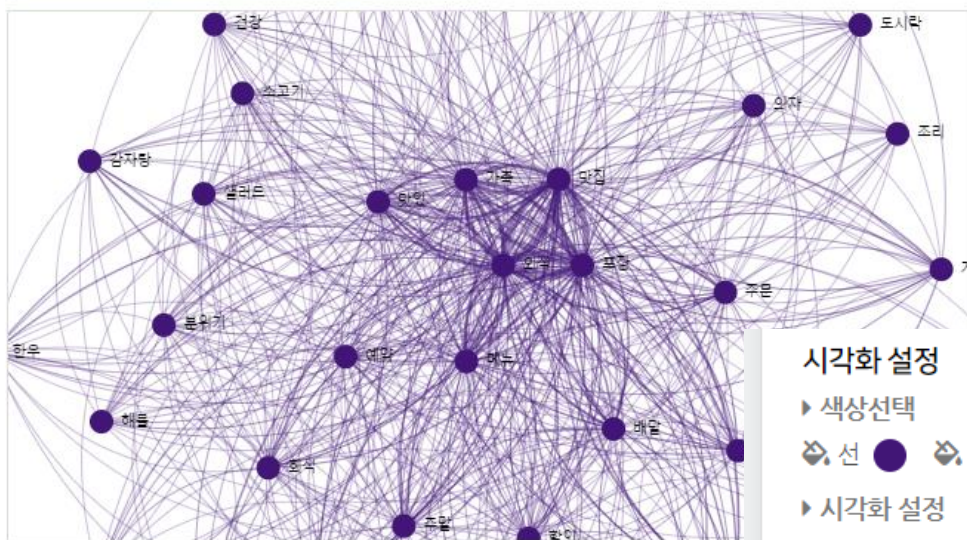
시각화 결과

크게보기

다운로드

매트릭스

N-gram



시각화 설정

▶ 색상선택

선    원    글자

▶ 시각화 설정

선 모양    글자 크기 10

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 매트릭스 분석은 2종의 시각화 결과를 제공합니다.

- 매트릭스, N-gram

④ [시각화 설정] 을 통해, 시각화의 색상과 모양 등을 변경할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.



1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 단어(Matrix Word)

1 **분석 알고리즘**

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다.  
최대 500개 단어를 선정할 있습니다. 200개 이상 선택 시, 파일로 업로드 해주세요.

예시파일

- ▶ 열단어 단어 선택하기 파일 업로드
- ▶ 행단어 단어 선택하기 파일 업로드

분석단어선택 (매트릭스 2-Mode 열단어 선택)

분석단어선택 (매트릭스 2-Mode 행단어 선택)

데이터명	생성일자	용량
교육부	2023-03-17	830.1 KB

선택단어수  확인 선택단어누적비율  % 다운로드 적용

상위 200개 까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/> 교육부	1546	5.448458149779736	5.448458149779736
<input type="checkbox"/> 교육	502	1.7691629955947137	7.21762114537445
<input type="checkbox"/> 학교	349	1.229955947136564	8.447577092511013
<input type="checkbox"/> 2	246	0.8669603524229075	9.31453744493392
<input type="checkbox"/> 있	210	0.7400881057268722	10.054625550660793
<input type="checkbox"/> 장관	187	0.6590308370044053	10.713656387665198

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는

번호

① **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② **[매트릭스 분석을 진행할 단어를 선택] : 단어 선택하기**

- 2-Mode 는 열과 행으로 구분하여 단어를 선택해주세요.

- [단어 선택하기] 를 클릭하면, 단어를 선택할 수 있는 새 창이 나타납니다.

③ 단어 선택을 완료한 후, **[적용]** 버튼을 눌러주세요.

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 단어(Matrix Word)

① [분석 알고리즘](#)

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다.  
최대 500개 단어를 선택할 있습니다. 200개 이상 선택 시, 파일로 업로드 해주세요.

[예시파일](#)

▶ 열단어

[단어 선택하기](#)

[파일 업로드](#)

열단어 파일을 업로드해 보세요.

[파일 업로드](#)

▶ 행단어

[단어 선택하기](#)

[파일 업로드](#)

행단어 파일을 업로드해 보세요.

[파일 업로드](#)

③ [적용](#)

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

번호

단어

빈도(건)

분석 결과가 존재하지 않습니다.

① **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② **[매트릭스 분석을 진행할 단어를 선택] : 파일 업로드**

- 2-Mode 는 열과 행으로 구분하여 단어를 선택해주세요.

- [파일 업로드] 를 클릭하면, 오른쪽에서 파일 업로드 창이 생성됩니다.

- 엑셀(xlsx), 메모장(txt)을 활용해 단어 목록을 작성해 업로드 해주세요.

※ 파일 업로드 시, 데이터에 등장하는 단어를 입력해주세요.

③ 단어 선택을 완료한 후, **[적용]** 버튼을 눌러주세요.

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

▶ 열단어

☰ 단어 선택하기

📁 파일 업로드

▶ 행단어

☰ 단어 선택하기

📁 파일 업로드

✓ 적용

전체 단어 중, 분석 적용된 단어의 데이터를 확인할 수 있습니다.

④ 📄 텍스트 다운로드 📊 엑셀 다운로드

⑤

번호	단어	빈도(건)
1	갈비	30
2	감자탕	14
3	부대찌개	15
4	간장	16
5	떡볶이	20
6	멸치	12
7	오리	11
8	계장	25
9	샐러드	11

④ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

⑤ 매트릭스 단어 결과로 매트릭스 분석에 선택한 단어의 빈도가 나타납니다.

데이터 정보 ↗
원문 정제

untitled

수집 날짜: 22.04.06   용량: 615 KB

시각화 결과
크게보기 다운로드

가로막대그래프

세로막대그래프

워드클라우드 3

**시각화 설정** 4

▶ 색상선택

▶ 빈도 ●

▶ 키워드 선정

선택해제

상위 50개

빈도

키워드

내림차순

오름차순

1~50

51~100

<input checked="" type="checkbox"/> 외식	335
<input checked="" type="checkbox"/> 포장	301
<input checked="" type="checkbox"/> 맛집	264

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 매트릭스 단어는 3종의 시각화 결과를 제공합니다.

- 가로막대 그래프, 세로막대 그래프, 워드클라우드

④ [시각화 설정] 을 통해, 시각화의 색상과 키워드를 변경할 수 있습니다.

- 빈도 : 막대 색상

- 빈도 또는 키워드(내림차순/오름차순) 기준으로 정렬하여 키워드를 선택할 수 있습니다.

※ 막대 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다.

1-Mode

2-Mode

매트릭스 단어 매트릭스 결과

### 매트릭스 결과(Matrix Result)

① 분석 알고리즘

바로선택하기/업로드를 통해 매트릭스 단어를 선정할 수 있습니다. (최대 500개 단어 선택 가능)  
200개 이상 선택 시, 파일로 업로드 해주세요.

매트릭스

엣지리스트

유클리디언 계수

②

코사인 계수

자카드 계수

상관 계수

전체 단어는 다운로드하여 확인할 수 있습니다.

③

텍스트 다운로드 엑셀 다운로드

메뉴	족발	찜	밥	갈비	게장	피자	닭	떡볶이	해물	간장	초밥	부대찌개	한우	감자탕	멸치	양념	반찬	샐러드	소고기	
포장	57	41	34	25	29	24	15	20	8	14	13	12	13	8	14	9	11	9	8	5
맛집	67	60	36	17	36	32	28	10	9	15	15	41	10	12	2	9	14	5	8	8
가족	23	16	16	4	7	4	4	7	2	10	6	10	4	8	7	3	1	2	2	5

① [분석 알고리즘]을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② 매트릭스 결과(Matrix Result)는 6종의 분석 결과를 매트릭스 형태로 제공합니다.

- 매트릭스, 엣지리스트, 유클리디언 계수, 코사인 계수, 자카드 계수, 상관계수

※ 분석 명칭을 클릭하면, 해당 분석결과를 하단에 나타냅니다.

③ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

데이터 정보

원문    경제

외식

수집 날짜 : 22.04.06    용량 : 615.00 KB

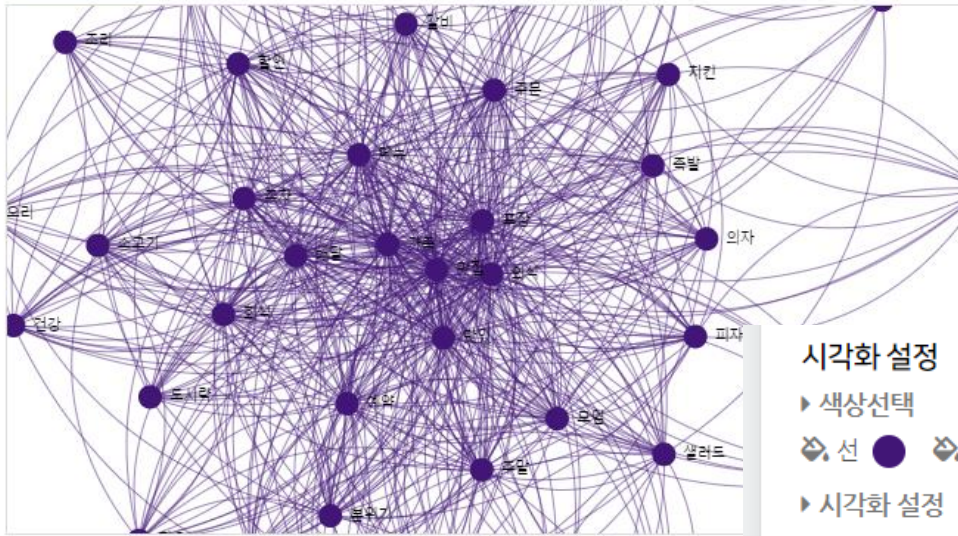
시각화 결과

크게보기

다운로드

매트릭스

N-gram



시각화 설정

시각화 설정

▶ 색상선택

선 ● 원 ● 글자 ●

▶ 시각화 설정

선 모양 1 2 3    글자 크기 10

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 매트릭스 분석은 2종의 시각화 결과를 제공합니다.

- 매트릭스, N-gram

④ [시각화 설정] 을 통해, 시각화의 색상과 모양 등을 변경할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## 네트워크 속성

**1** 분석 알고리즘

네트워크 전체에 대한 구조적 기술 통계량을 분석합니다. 네트워크의 노트들이 상호 간에 얼마나 연결이 잘되어 있는가를 관련 지수로 측정합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

**2** 텍스트 다운로드 엑셀 다운로드

<b>3</b>	Network Measures	Value
	Nodes	40
	Total Edges	336
	Diameter	2
	Degree Centralization	0.59919
	Closeness Centralization	0.71444
	Betweenness Centralization	0.00011
	Connected Components	1
	Overall Reciprocity	0

**1** [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** 네트워크 속성 분석결과는 1-Mode 매트릭스에 대한 네트워크의 구조적 속성 지표 8종을 산출합니다.

1	Nodes	노드의 총 개수	5	Closeness Centralization	근접 집중도
2	Total Edges	엣지의 총 개수	6	Betweenness Centralization	매개 집중도
3	Diameter	직경	7	Connected Components	연결된 컴포넌트 수
4	Degree Centralization	연결정도 집중도	8	Overall Reciprocity	상호성

※ 네트워크 속성은 1-Mode 매트릭스 결과를 진행하여야 나타나며, 시각화결과가 없습니다.

네트워크 속성

**중심성**

에고 네트워크

최단경로

CONCOR

클러스터링

## 중심성

**1** [분석 알고리즘](#)

네트워크에서 특정 노드(Node)가 갖고 있는 상대적 중요성을 수치화한 값을 계산합니다.  
중심성 분석을 통해 구조적 위치의 관점에서 보다 더 영향력 있는(의미있는) 노드를 파악할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

**2** [텍스트 다운로드](#) [엑셀 다운로드](#)

<b>3</b> 단어	연결정도 중심성	근접 중심성	매개 중심성	아이겐벡터 중심성	페이지 랭크	클러스터링 계수
외식	1.0000000	1.0000000	0.0987763	0.2744513	0.1443191	0.4008097
포장	1.0000000	1.0000000	0.0987763	0.2744513	0.1326622	0.4008097
맛집	1.0000000	1.0000000	0.0987763	0.2744513	0.1403977	0.4008097
가족	0.9487179	0.9512195	0.0755708	0.2685696	0.0644874	0.4294294
메뉴	0.7435897	0.7959184	0.0349259	0.2267818	0.0401853	0.5098522
맛있	0.7692308	0.8125000	0.0440022	0.2274611	0.0403728	0.4804598
배달	0.6410256	0.7358491	0.0219752	0.2039866	0.0323150	0.5766667
족발	0.3846154	0.6190476	0.0051365	0.1416613	0.0287037	0.7333333
주문	0.6153846	0.7222222	0.0216081	0.1946801	0.0184932	0.5688406
갈비	0.4358974	0.6393443	0.0047601	0.1591484	0.0195032	0.7573529

**1** **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

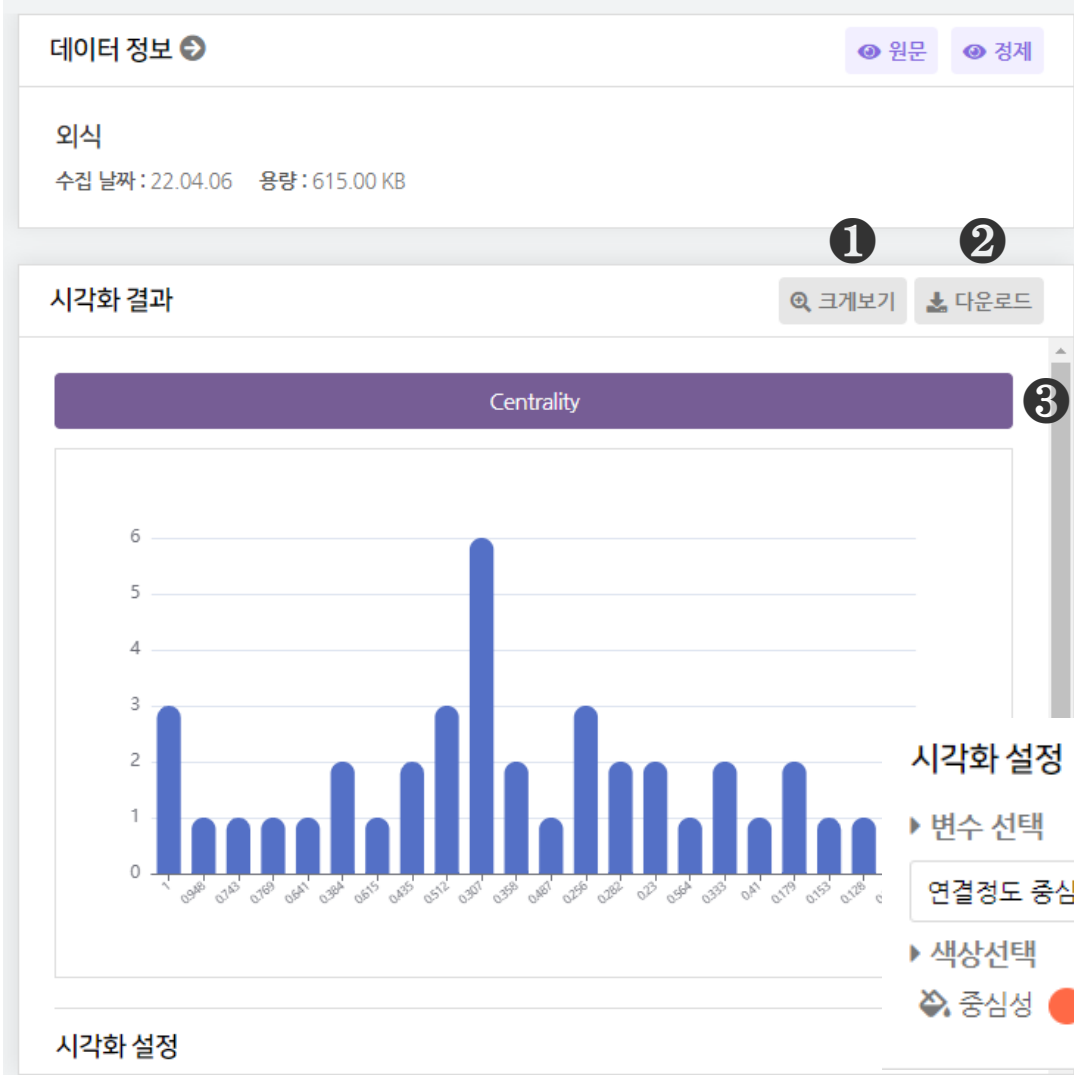
**2** **[텍스트 다운로드], [엑셀 다운로드]**를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** 중심성 분석결과로 노드(단어)별 네트워크 내 중심성 척도 6종을 산출합니다.

- 연결정도 중심성, 근접 중심성, 매개 중심성, 아이겐벡터 중심성, 페이지 랭크, 클러스터링 계수

※ 중심성은 1-Mode 매트릭스 결과를 진행하여야 나타납니다.





① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 중심성은 1종의 시각화 결과를 제공합니다.

- Centrality(x=중심성 값, y=중심성 값의 빈도)

④ [시각화 설정] 을 통해, 시각화의 색상과 변수를 변경할 수 있습니다.

- 변수 선택 : 연결정도 중심성, 근접 중심성, 매개 중심성, 아이겐벡터 중심성, 페이지 랭크, 클러스터링 계수 중 선택할 수 있습니다.

※ 막대 위에 마우스를 올리면 데이터 값(빈도)를 확인할 수 있습니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## 에고 네트워크

1 분석 알고리즘

에고 네트워크는 한 노드를 에고(Ego)에 위치시키고, 에고와 다른 노드(타자, Alter)와의 연결을 표현한 네트워크를 의미합니다. 텍스트은 전체 네트워크에서 특정노드(에고)를 대상으로 하는 에고 네트워크를 추출하며, 에고와 연결된 타자노드들 간의 연결 관계를 분석합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

3 단어	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
멸치	5	15	20	0.75	5	0
창업	6	20	30	0.66666	10	0.00021
다이어트	7	27	42	0.64285	15	0.00008
여행	7	25	42	0.59523	17	0.00044
오리	8	34	56	0.60714	22	0.00012
초밥	9	41	72	0.56944	31	0.00016
카페	9	40	72	0.55555	32	0.00034
해물	10	50	90	0.55555	40	0.00012
감자탕	10	50	90	0.55555	40	0.00012
조리	10	50	90	0.55555	40	0.00011

1 [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

3 에고 네트워크 분석결과로 노드(단어)별 에고 네트워크 지표 6종을 산출합니다.

1 Size	에고 네트워크 크기	4 Ego-Density	에고 네트워크 밀도
2 Ties	직접 연결된 엣지 수	5 Broker	중개자
3 Pairs	전체 노드 쌍의 수	6 Ego Betweenness	에고 매개성

※ 중심성은 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

데이터 정보
원문    경제

untitled

수집 날짜 : 22.04.06    용량 : 615 KB

시각화 결과
크게보기    다운로드

에고네트워크

**시각화 설정**

중심 단어 입력

중심단어를 입력해 주세요 적용

색상선택

배경색 상 중 중    글자색 ●

키워드 선정

선택해제		상위 50개	
빈도	키워드	내림차순	오름차순
1~50			
<input checked="" type="checkbox"/>	맛있	0	
<input checked="" type="checkbox"/>	가족	0	
<input checked="" type="checkbox"/>	포장	0	
<input checked="" type="checkbox"/>	맛집	0	
51~100			

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 에고 네트워크는 1종의 시각화 결과를 제공합니다.

④ [시각화 설정] 을 통해, 시각화의 색상 변경과 중심 단어 입력, 키워드를 변경할 수 있습니다.

- 중심 단어 입력 : 에고 네트워크에서 에고에 해당하는 단어를 입력하고, 적용 버튼을 누릅니다.

- 키워드 선정 : 에고 네트워크의 타자 노드들을 선택/해제할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

네트워크 속성

중심성

예고 네트워크

최단경로

CONCOR

클러스터링

## 최단경로

**1** [분석 알고리즘](#)

경로(Path)는 연결된 노드들을 거쳐 한 노드에서 다른 노드까지 가는 길을 의미합니다. 최단 경로(Shortest path)는 네트워크 내 가장 길이가 짧은 경로를 말합니다.

▶ 경로시작점

**2**

▶ 경로도착점

**3**

✓ 적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

경로시작점

경로도착점

최단경로 수

경로

분석할 경로의 단어를 입력해주세요.

**1** **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** **[경로시작점], [경로도착점]** 을 입력해주세요.

- 경로시작점(Source) : 네트워크 내에서 경로를 확인하고자 하는 최초 시작 단어를 입력

- 경로도착점(Target) : 네트워크 내에서 경로를 확인하고자 하는 최종 도착 단어를 입력

※ 경로시작점과 경로도착점은 네트워크 내 존재하는 노드(단어)이어야 합니다.

**3** **[적용]** 버튼을 눌러, 분석을 진행합니다.

※ 최단경로는 1-Mode 매트릭스 결과를 진행하여야 분석 가능합니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## 최단경로

1 분석 알고리즘

경로(Path)는 연결된 노드들을 거쳐 한 노드에서 다른 노드까지 가는 길을 의미합니다. 최단 경로(Shortest path)는 네트워크 내 가장 길이가 짧은 경로를 말합니다.

▶ 경로시작점

창업

▶ 경로도착점

갈비

✓ 적용

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

4

텍스트 다운로드 엑셀 다운로드

5

경로시작점

경로도착점

최단경로 수

경로

창업

갈비

3

창업,외식,갈비

창업

갈비

3

창업,포장,갈비

창업

갈비

3

창업,맛집,갈비

4 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

5 최단경로 분석결과로 입력한 경로시작점에서 경로도착점까지 최단경로의 수와 경로를 산출합니다.

- 최단경로 수 : 경로시작점에서 경로도착점까지 도달하는 가장 짧은 경로의 경우의 수

- 경로 : 경로시작점에서 경로도착점까지 도달하는 노드의 경로를 나타냅니다.

※ 최단경로는 1-Mode 매트릭스 결과를 진행하여야 분석 가능합니다.

데이터 정보 ↗
원문 정제

---

untitled  
수집 날짜: 22.04.06 용량: 615 KB

---

시각화 결과
크게보기 다운로드

---

네트워크 차트

3

시각화 설정 4

- ▶ 글자 설정
- ▶ 색상 ● Tt 글자 크기 10
- ▶ 노드 설정
- 크기 기본 1 2 3
- ▶ 엣지 설정
- ▶ 색상 ● ↔ 굵기 1 2 3

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 최단경로는 1종의 시각화 결과를 제공합니다.

④ [시각화 설정] 을 통해, 시각화의 글자, 노드, 엣지의 설정을 변경할 수 있습니다.

- 글자 설정 : 글자 색상과 크기를 변경합니다.
- 노드 설정 : 노드 크기를 변경합니다.
- 엣지 설정 : 엣지의 색상과 굵기를 변경합니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## CONCOR

1 [분석 알고리즘](#)

분석은 동시에 출현하는 단어 사이의 관계를 분석하여 구조적 등위의 위치에 있는 단어들을 군집(블록)하는 분석 기법입니다.

▶ 군집 수

4

2

⚙️ 속성값 설정 [다운로드 시 전체를 확인할 수 있습니다.](#)

3

[텍스트 다운로드](#) [엑셀 다운로드](#)

4

분석모델	단어	그룹
CONCOR(concor)	반찬	1
CONCOR(concor)	샐러드	1
CONCOR(concor)	소고기	1
CONCOR(concor)	오리	1
CONCOR(concor)	칼국수	1
CONCOR(concor)	치킨	1
CONCOR(concor)	여행	1
CONCOR(concor)	의자	1
CONCOR(concor)	도시락	1

1 **[분석 알고리즘]**을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 **[군집 수]**를 선택합니다.

3 **[텍스트 다운로드]**, **[엑셀 다운로드]** 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

4 CONCOR 분석결과로 단어와 그 단어가 속한 그룹을 산출합니다.

※ CONCOR 는 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## CONCOR

[분석 알고리즘](#)

분석은 동시에 출현하는 단어 사이의 관계를 분석하여 구조적 등위의 위치에 있는 단어들을 군집(블록)하는 분석 기법입니다.

▶ 군집 수

4

### 속성값 설정

속성값 설정

5

데이터 시 전체를 확인할 수

분석모드

- 단어빈도
- 연결중심성
- 근접중심성
- 매개중심성
- 위세중심성

속성값은 텍스트마이닝, 중심성 분석이 완료된 경우 적용가능합니다.

선택 적용

텍스트 다운로드

엑셀 다운로드

그룹

CONCOR(c		1
CONCOR(c		1
CONCOR(c		1
CONCOR(c		1
CONCOR(c		1
CONCOR(c		1
CONCOR(concor)	치킨	1
CONCOR(concor)	여행	1
CONCOR(concor)	의자	1
CONCOR(concor)	도시락	1

⑤ [속성값 설정] 을 통해 단어빈도, 중심성 분석 결과를 불러올 수 있습니다.

불러올 속성값을 선택한 후, [선택 적용] 버튼을 눌러주세요.

※ 속성값은 단어분석, 중심성 분석이 완료된 경우 적용 가능합니다.

※ CONCOR 는 1-Mode 매트릭스 결과를 진행하여야 나타납니다.



네트워크 속성

중심성

에고 네트워크

최단경로

**CONCOR**

클러스터링

## CONCOR

[분석 알고리즘](#)

분석은 동시에 출현하는 단어 사이의 관계를 분석하여 구조적 등위의 위치에 있는 단어들을 군집(블록)하는 분석 기법입니다.

▶ 군집 수

4

⚙️ 속성값 설정

다운로드 시 전체를 확인할 수 있습니다.

📄 텍스트 다운로드

📊 엑셀 다운로드

분석모델	단어	그룹	단어빈도
CONCOR(concor)	반찬	1	12
CONCOR(concor)	샐러드	1	11
CONCOR(concor)	소고기	1	11
CONCOR(concor)	오리	1	11
CONCOR(concor)	6 칼국수	1	11
CONCOR(concor)	치킨	1	10
CONCOR(concor)	여행	1	10
CONCOR(concor)	의자	1	10
CONCOR(concor)	도시락	1	10

⑥ CONCOR 분석결과에 **[속성값 설정]** 을 통해 불러온 속성값을 추가하여 산출합니다.

불러온 속성값은 **[시각화 설정]** 에서 노드의 크기에 적용할 수 있습니다.

※ CONCOR 는 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

데이터 정보

원문

정제

untitled

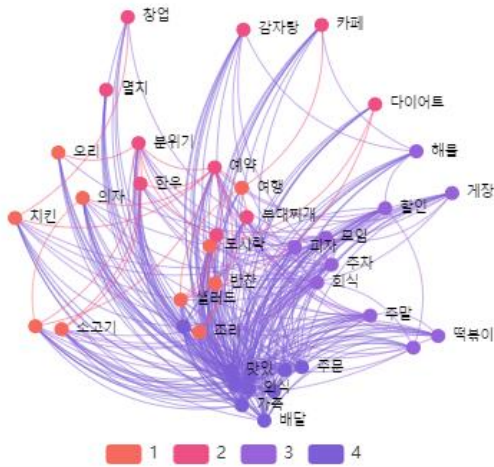
수집 날짜: 22.04.06 용량: 615 KB

시각화 결과

크게보기

다운로드

네트워크 차트



시각화 설정

시각화 설정

글자 설정

색상  글자 크기 10

노드 설정

크기 기본 1 2 3

엣지 설정

색상  굵기 1 2 3

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ CONCOR 는 1종의 시각화 결과(네트워크 차트)를 제공합니다.

④ [시각화 설정] 을 통해, 글자, 노드, 엣지의 색상과 크기 등을 변경할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## 클러스터링

1 ⓘ 분석 알고리즘

어떠한 형태의 소규모 그룹이 존재하는지 파악하기 위한 계층적 군집 분석(Hierarchical Clustering)입니다.  
 각 노드가 다른 노드들과 맺는 관계의 패턴을 계산하여, 유사한 관계패턴을 갖는 노드들이 동일한 군집으로 분류됩니다.

▶ 분석모델 선택

루베인(Louvain)

2

\* 속성값 설정 다운로드 시 전체를 확인할 수 있습니다.

3 📄 텍스트 다운로드 📄 엑셀 다운로드

4	분석모델	단어	그룹
	루베인(Louvain)	떡볶이	1
	루베인(Louvain)	예약	1
	루베인(Louvain)	치킨	1
	루베인(Louvain)	외식	2
	루베인(Louvain)	포장	2
	루베인(Louvain)	맛집	2
	루베인(Louvain)	가족	2
	루베인(Louvain)	메뉴	2
	루베인(Louvain)	배달	2

1 [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 [분석모델 선택] 을 선택합니다.

- 클러스터링 분석은 루베인(Louvain), 레이든(Leiden),  
 GN(girvan\_newman), CNM(clauset\_newman\_moore) 알고리즘 4종을 제공합니다.

3 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과  
 파일이 다운로드 됩니다.

4 클러스터링 분석결과로 단어와 그 단어가 속한 그룹을 산출합니다.

※ 클러스터링 은 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

## 클러스터링

[분석 알고리즘](#)

어떠한 형태의 소규모 그룹이 존재하는지 파악하기 위한 계층적 군집 분석(Hierarchical Clustering)입니다. 각 노드가 다른 노드들과 맺는 관계의 패턴을 계산하여, 유사한 관계패턴을 갖는 노드들이 동일한 군집으로 분류됩니다.

▶ 분석모델 선택

루베인(Louvain)

속성값 설정

5

속성값 설정

텍스트 다운로드 엑셀 다운로드

분석모델	속성값	그룹
루베인(Louvain)	단어빈도	1
루베인(Louvain)	연결중심성	1
루베인(Louvain)	근접중심성	1
루베인(Louvain)	매개중심성	1
루베인(Louvain)	위세중심성	1
루베인(Louvain)	속성값은 텍스트마이닝, 중심성 분석이 완료된 경우 적용가능합니다.	1
루베인(Louvain)		2
루베인(Louvain)		2
루베인(Louvain)	맛집	2
루베인(Louvain)	가족	2
루베인(Louvain)	배달	2
루베인(Louvain)	회식	2

5 [속성값 설정] 을 통해 단어빈도, 중심성 분석 결과를 불러올 수 있습니다.  
 불러올 속성값을 선택한 후, [선택 적용] 버튼을 눌러주세요.

※ 속성값은 단어분석, 중심성 분석이 완료된 경우 적용 가능합니다.

※ 클러스터링 은 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

네트워크 속성

중심성

예고 네트워크

최단경로

CONCOR

클러스터링

## 클러스터링

① 분석 알고리즘

어떠한 형태의 소규모 그룹이 존재하는지 파악하기 위한 계층적 군집 분석(Hierarchical Clustering)입니다.  
 각 노드가 다른 노드들과 맺는 관계의 패턴을 계산하여, 유사한 관계패턴을 갖는 노드들이 동일한 군집으로 분류됩니다.

▶ 분석모델 선택

루베인(Louvain)

⚙ 속성값 설정

다운로드 시 전체를 확인할 수 있습니다.

📄 텍스트 다운로드

📄 엑셀 다운로드

분석모델	단어	그룹	단어빈도
루베인(Louvain)	떡볶이	1	20
루베인(Louvain)	예약	1	15
루베인(Louvain)	치킨	1	10
루베인(Louvain)	맛있	2	65
루베인(Louvain)	한우	2	15
루베인(Louvain)	다이어트	2	13
루베인(Louvain)	카페	2	12
루베인(Louvain)	반찬	2	12
루베인(Louvain)	칼국수	2	11

⑥ 클러스터링 분석결과에 [속성값 설정] 을 통해 불러온 속성값을 추가하여 산출합니다.

- 불러온 속성값은 [시각화 설정] 에서 노드의 크기에 적용할 수 있습니다.

※ 클러스터링 은 1-Mode 매트릭스 결과를 진행하여야 나타납니다.

데이터 정보 원문 경제

외식  
수집 날짜: 22.04.06 용량: 615.00 KB

시각화 결과 크게보기 다운로드

네트워크 차트

시각화 설정

글자 설정  
색상 ● 글자 크기 10

노드 설정  
크기 기본 1 2 3

엣지 설정  
색상 ● 굵기 1 2 3

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 클러스터링은 1종의 시각화 결과(네트워크 차트)를 제공합니다.

④ [시각화 설정]을 통해, 시각화의 글자, 노드, 엣지 설정을 변경할 수 있습니다.

- 노드 크기 : 속성값 설정을 통해 속성값을 불러온 경우,  
속성값을 노드 크기에 적용할 수 있습니다.

※ 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

LDA

토픽 성능 토픽 단어 토픽 문서

토픽 성능

① 분석 알고리즘

토픽 분석으로 유사한 의미를 가진 키워드들을 주제별(Topic)로 묶어 텍스트의 잠재적인 의미 구조를 파악합니다.  
LDA(Latent Dirichlet Allocation)은 토픽 모델링의 대표적인 알고리즘입니다

▶ Alpha  ▶ Beta  ▶ Iterations

▶ 토픽 수  개 ▶ 토픽 단어 수  개 ▶ 랜덤 값

③

전체 목록은 [다운로드하여 확인](#)할 수 있습니다.

[텍스트 다운로드](#) [엑셀 다운로드](#)

토픽수

Perplexity Score

Coherence Score

분석 결과가 존재하지 않습니다.

① [분석 알고리즘]을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② LDA 분석을 진행하기 위해 다음의 분석 조건 값을 설정해주세요.

- [Alpha], [Beta] : 0.01~1.00 범위의 수
- [Iterations] : 50~2000 범위의 수
- [토픽 수], [토픽 단어의 수] : 1~100 범위의 수
- [랜덤 값] **사용** 시, 무작위 할당을 진행합니다(재현성이 떨어짐).

분석결과의 재현성을 확보하고 싶은 경우 **사용안함**을 선택해주세요.

③ 분석 조건 설정을 완료한 후, [적용] 버튼을 눌러주세요.

LDA

토픽 성능   토픽 단어   토픽 문서

▶ 토픽 수  개   ▶ 토픽 단어 수  개   ▶ 랜덤 값

전체 목록은 다운로드하여 확인할 수 있습니다.

④

⑤ 토픽수	Perplexity Score	Coherence Score
5	-10.885252788726419	-2.009856040267465
6	-11.531940142252056	-2.0997215204939956
7	-12.135766372113485	-2.0070535821158964
8	-12.719211041211436	-1.9919733450753034
9	-13.27861988881631	-1.9606805018974915
10	-13.811892040174905	-1.954244690510734
2	-8.770335749397903	-1.9340845181622064
3	-9.511733457451871	-2.0349972119504813
4	-10.223334132493443	-2.098426372037771

④ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석결과 파일이 다운로드 됩니다.

⑤ 토픽성능 분석결과로 다음의 지표를 산출합니다.

1) Perplexity Score : 토픽모델의 예측 성능 지표로, 값이 작을 수록 토픽모델이 주어진 텍스트 데이터에 대해 학습을 효과적으로 수행한 것으로 판단합니다.

2) Coherence Score : 토픽모델 내 주제의 일관성 지표로, 값이 높을 수록 토픽 내 유사한 단어가 많이 등장함을 의미합니다.

- 일반적으로 Perplexity 값이 낮으면서, Coherence 값이 높을 때, 최적의 토픽 개수를 갖는다고 볼 수 있습니다.

Tip. 토픽성능 분석결과를 통해 최적의 토픽 개수를 확인한 후, LDA 조건에 해당 토픽 개수로 재설정하여 분석을 진행해보세요.



데이터 정보

원문

정제

외식

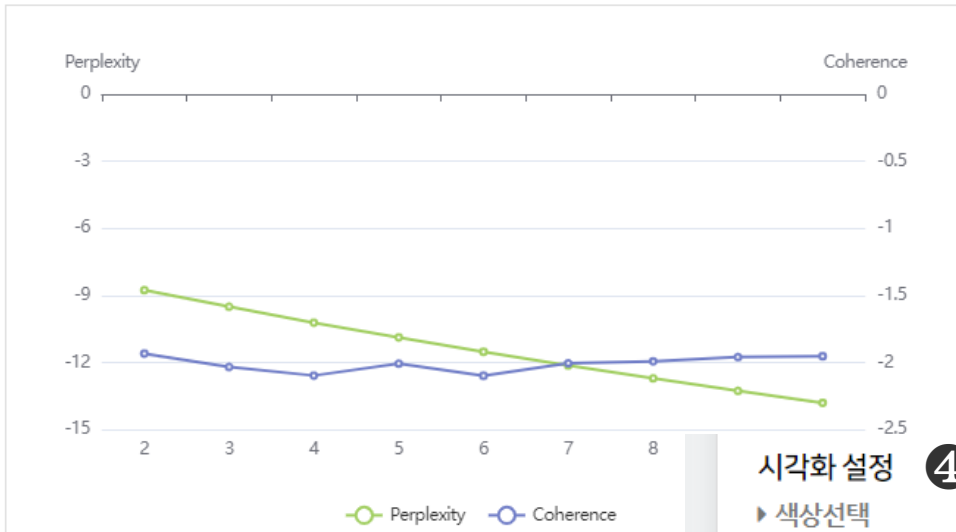
수집 날짜 : 22.04.06    용량 : 615.00 KB

시각화 결과

크게보기

다운로드

토픽 성능



시각화 설정

색상선택

시각화 설정

색상선택

Perplexity Coherence

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 토픽 성능은 1종의 시각화(토픽성능)결과를 제공합니다.

④ [시각화 설정] 을 통해, 시각화의 색상을 변경할 수 있습니다.

LDA

토픽 성능 토픽 단어 토픽 문서

토픽 단어

1 분석 알고리즘

토픽 분석으로 유사한 의미를 가진 키워드들을 주제별(Topic)로 묶어 텍스트의 잠재적인 의미 구조를 파악합니다. LDA(Latent Dirichlet Allocation)은 토픽 모델링의 대표적인 알고리즘입니다

전체 목록은 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

3	토픽	단어	단어 토픽확률
	1	포장	0.03
	1	외식	0.03
	1	맛집	0.025
	1	먹	0.023
	1	집	0.015
	1	있	0.014
	1	총	0.013
	1	하	0.011
	1	가족	0.009

1 [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석결과 파일이 다운로드 됩니다.

3 토픽단어 분석결과로 추출된 토픽과 단어의 토픽확률(%) 산출합니다.  
- 단어 토픽확률은 토픽에 해당 단어가 할당될 확률을 의미합니다.

데이터 정보

원문    정제

외식

수집 날짜 : 22.04.06    용량 : 615 KB

시각화 결과

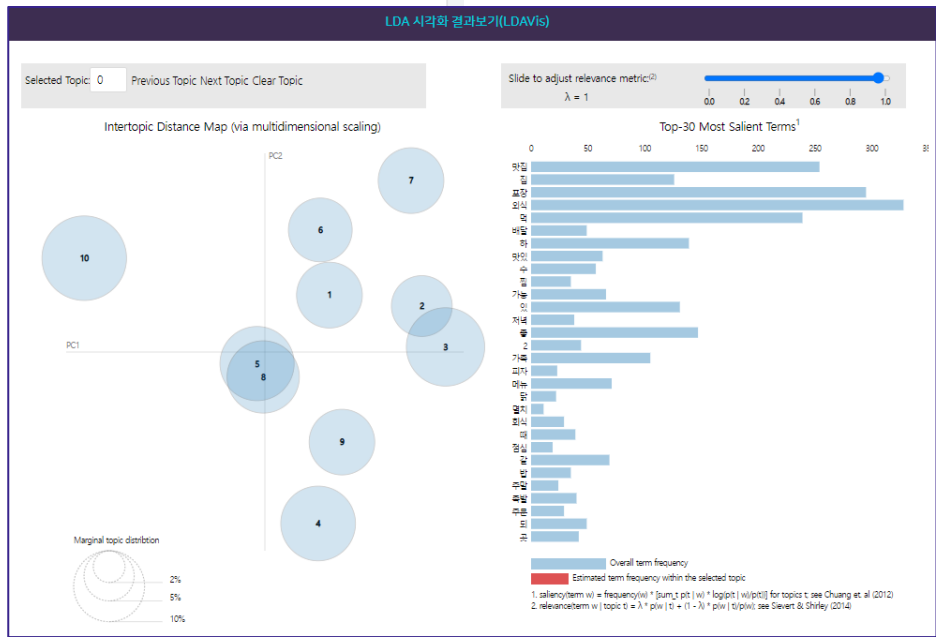
크게보기

다운로드

LDAs

워드클라우드

LDAs 결과보기



① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 토픽 단어 분석은 2종의 시각화 결과를 제공합니다.

- LDAs, 워드클라우드
- LDAs 는 [LDAs 결과보기]를 클릭하여 확인합니다.

데이터 정보 원문 정제

외식  
수집 날짜: 22.04.06 용량: 615 KB

시각화 결과 크게보기 다운로드

LDAVis 워드클라우드

시각화 설정

① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 토픽 단어 분석은 2종의 시각화 결과를 제공합니다.

- LDAVis, 워드클라우드
- 워드클라우드 는 토픽별로 할당된 단어의 시각화입니다.

LDA

토픽 성능 토픽 단어 토픽 문서

토픽 문서

1 분석 알고리즘

토픽 분석으로 유사한 의미를 가진 키워드들을 주제별(Topic)로 묶어 텍스트의 잠재적인 의미 구조를 파악합니다. LDA(Latent Dirichlet Allocation)은 토픽 모델링의 대표적인 알고리즘입니다

전체 목록은 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

3	토픽	문서 토픽확률	원문
	8	0.5924213	10 양산 천태산 테마 산행 안내 덕 천동 외식 1 가 세 본일레븐 편의점 앞 08 00 김해 시청 맞은편 탑 송지 차량 출발 5 전 도착 산행 코스 산 소개 1 산행 코스
	7	0.7373032	용인시 주택 경매 경기 용인시 수지구 신봉동 단지 임 야 농경지 등 혼재 신봉동 외식 타운 등 근린 시설 생 활 편리 마을버스 지상 물건 담장 수목 바닥 포장 화단 옹벽 등 토지 부합물
	8	0.28231585	3 마시 잔 수 있선 포장 약 규정 것 것 인생 안 약 도모 유익 주 연고

1 [분석 알고리즘] 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석결과 파일이 다운로드 됩니다.

3 토픽 문서 분석결과로 추출된 토픽에 할당되는 원문과 해당 원문의 토픽확률(%)을 산출합니다.

※ LDA 토픽 문서는 시각화결과가 없습니다.

문서분류

감성어휘

분류 성능   문서 감성 종합   긍정 문서   중립 문서   부정 문서

### 분류 성능

① [분석 알고리즘](#)

텍스트 문서에 대한 감성을 분석합니다.  
 베이지안 분류기(Bayes Classifier)를 활용해 기계학습 기법의 감성분석을 진행합니다.

[예시파일](#)

예시파일을 참고하여 최소 100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링하고 전체 데이터 중 80%를 학습데이터로 20%를 테스트 데이터로 각각 다른 파일을 업로드 해주세요. (비율을 비슷하게 업로드 할 수록 정확한 결과가 나옵니다.)

▶ 학습데이터

▶ 테스트데이터

✓ 적용 ③

[텍스트 다운로드](#) [엑셀 다운로드](#)

구분	정확도(%)
분석 결과가 존재하지 않습니다.	

① **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

② 감성 문서분류 분석을 진행하기 위해 **학습데이터와 테스트데이터를 업로드해 주세요.**

- 학습 데이터 : 전체 데이터를 분류하기 위한 기준이 되는 데이터로, 각 극성의 비율을 비슷하게 태깅 하여 업로드해 주세요.
- 테스트데이터 : 학습데이터를 기반으로 만들어진 분류기의 성능을 확인하기 위한 테스트 데이터를 업로드해 주세요.

③ 학습데이터와 테스트데이터 업로드 완료한 후, **[적용]** 버튼을 눌러주세요.

문서분류

감성어휘

분류 성능    문서 감성 종합    긍정 문서    중립 문서    부정 문서

### 분류 성능

[분석 알고리즘](#)

텍스트 문서에 대한 감성을 분석합니다.  
 베이지안 분류기(Bayes Classifier)를 활용해 기계학습 기법의 감성분석을 진행합니다.

[예시파일](#)

예시파일을 참고하여 최소100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링하고 전체 데이터 중 80%를 학습데이터로 20%를 테스트 데이터로 각각 다른 파일을 업로드해주세요. (비율을 비슷하게 업로드 할 수록 정확한 결과가 나옵니다.)

▶ 학습데이터

학습데이터를 업로드해 보세요.



▶ 테스트데이터

테스트데이터를 업로드해 보세요.



✓ 적용

4

[텍스트 다운로드](#)   [엑셀 다운로드](#)

5

구분

정확도(%)

전체정확도(overall accuracy)

49.02

긍정

68.18

중립

50.0

부정

9.09

④ [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

⑤ 감성 문서분류 성능결과로 긍정문서, 중립문서, 부정문서의 분류정확도를 산출합니다.

구분		실제 정답			극성별 정확도
		긍정	중립	부정	
분류 결과	긍정	A	B	C	$A/(A+B+C)*100$
	중립	D	E	F	$E/(D+E+F)*100$
	부정	G	H	I	$I/(G+H+I)*100$
Overall Accuracy(전체 정확도)		$(A+E+I)/(A+B+C+D+E+F+G+H+I)*100$			

데이터 정보 →
원문    정제

---

외식  
수집 날짜: 22.04.06    용량: 615 KB

시각화 결과

1 
2

Confusion Matrix

	72	0	1
긍정			
중립	12	40	16
부정	2	0	26
	정	립	부

시각화 설정 3

▶ 색상선택

배경 색상 ●    글자 색상 ●

- ① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.
- ② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.
- ③ 감성 문서분류 성능은 1종의 시각화(Confusion Matrix)결과를 제공합니다.
- ④ [시각화 설정] 을 통해, 시각화의 색상을 변경할 수 있습니다.



문서분류

감성어휘

분류 성능   **문서 감성 종합**   긍정 문서   중립 문서   부정 문서

### 문서 감성 종합

**1** [분석 알고리즘](#)

텍스트 문서에 대한 감성을 분석합니다.  
 베이지안 분류기(Bayes Classifier)를 활용해 기계학습 기법의 감성분석을 진행합니다.

다운로드 시 전체를 확인할 수 있습니다.

**2** [텍스트 다운로드](#) [엑셀 다운로드](#)

구분	빈도(건)	비율(%)
전체	0	100.0
<b>긍정</b>	<b>3</b>	48.3
중립	429	42.9
부정	88	8.8

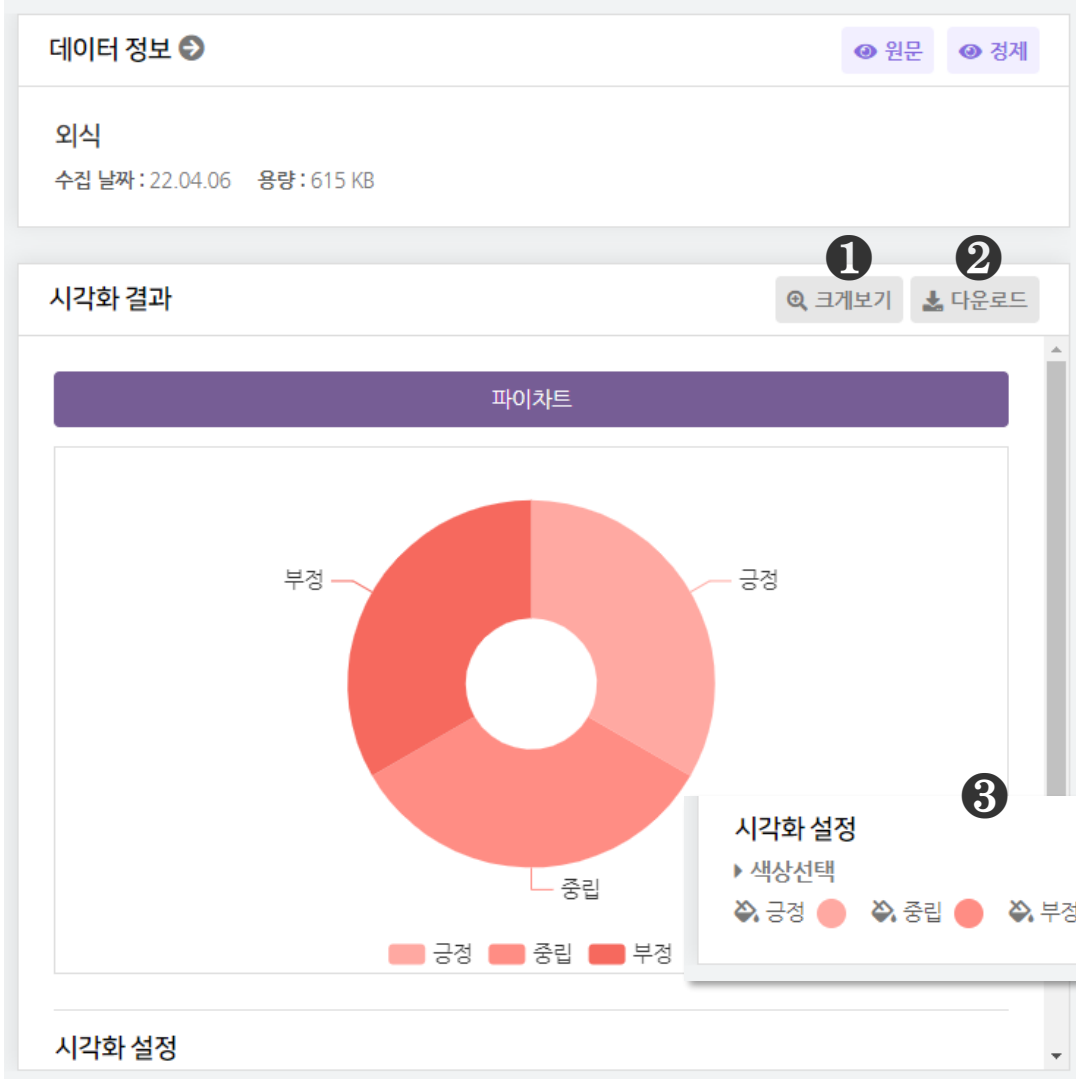
분석문장	결과
교육부 "2학기 전면등교 변함없다"...서울·경기 학원강사... 갑작스런 코로나 확진자 증가세에 오늘 교육부에서도 조치들을 내놓았네요.. 학원... 첨부파일 [교육부+07 - 08 (목) +10시보도자료] +학교+학원+방역+강화+조치사항..	중립
교무 학사개입, 신입생 허위 모집 의혹 사실로...교육부, 전총건... '고발' 예고 교육부가 28일 공개한 김포대학 종합감사 처분서에는 전총건 이사장이 교무학사에 어떤 방식으로 부당 개입했는지 상세히 나타나있다.(자료=교육부)..	중립

**1** [\[분석 알고리즘\]](#)을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [\[텍스트 다운로드\]](#), [\[엑셀 다운로드\]](#) 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** 문서 감성 종합 분석결과로 전체/긍정/중립/부정 문서의 빈도와 비율(%)에 대한 요약 결과를 산출합니다.

**4** 문서 감성 종합 분석결과로 분류된 분석문장을 표로 제시합니다.



- ① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.
- ② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.
- ③ 문서 감성 종합 분석은 1종의 시각화 결과(파이차트)를 제공합니다.
- ④ [시각화 설정] 을 통해, 시각화의 색상을 변경할 수 있습니다.

문서분류

감성어휘

분류 성능   문서 감성 종합   **긍정 문서**   중립 문서   부정 문서

### 긍정 문서

**1** 분석 알고리즘

텍스트 문서에 대한 감성을 분석합니다.  
 베이지안 분류기(Bayes Classifier)를 활용해 기계학습 기법의 감성분석을 진행합니다.

다운로드 시 전체를 확인할 수 있습니다.

**2** 텍스트 다운로드   엑셀 다운로드

구분	빈도(건)	비율(%)
전체	<b>3</b>	100.0
<b>긍정</b>	483	48.3

분석문장	결과
재외교육지원청 (feat. 교육부, 재외한국학교) 물론 교육부에서도 지원금을 보내주지만, 학교 재정의 많은 부분이 학생들의... #재외한국학교 #재외교육기관 #교사해외근무 #재외등포재단 #교육부에게바란다..	긍정
모션과 IoT게임러닝시스템으로 교육부장관상까지 특히 최소기능제품(MVP) 구구단부터 론칭하며 시장으로부터 큰 호응을 얻었으며, 이를 기반으로 디바이스 및 시스템 고도화를 이루어 교육부장관상 및 특허기술상..	긍정
{압구정소아정신과} 교육부 위탁터 위촉 [밝은미래정신과] 교육부에서는 원격 화상 자문 시범사업을 진행하고 있습니다. 환자를 대상으로 한... 마음으로만 감사함을 늘 가지고 있던 중, 교육부 위탁터 시범사업의 위탁터로..	긍정

**1** [분석 알고리즘]을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

**2** [텍스트 다운로드], [엑셀 다운로드]를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**3** 긍정/중립/부정 문서 분석결과로 추출된 문서 빈도와 비율(%)을 산출합니다.

**4** 긍정/중립/부정 문서 분석결과로 분류된 분석문장을 표로 제시합니다.

※ 긍정/중립/부정 문서는 시각화결과가 없습니다.

문서분류

감성어휘

## 감성 어휘 분석

1 **분석 알고리즘**

TEXTOM에서 제작한 감성 어휘 사전을 바탕으로 단어를 분류하여, 빈도와 감성강도를 계산합니다.

▶ 분석데이터 선택

원문데이터

정제데이터

2

▶ 분석데이터 확인

다운로드(Exel)

다운로드(txt)

3

✓ 적용

4

다운로드 시 전체를 확인할 수 있습니다.

텍스트 다운로드 엑셀 다운로드

구분	빈도(건)	감성강도비율(%)	빈도비율(%)
긍정	0/0	0/100.0	0/100.0
부정	0/0	0/100.0	0/100.0

감정분류	빈도(건)	감정강도	빈도*감성강도	빈도비율(%)
긍정 키워드				
부정 키워드				
세부 감정 키워드				

데이터가 존재하지 않습니다.

1 **[분석 알고리즘]** 을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.

2 감성어휘 분석을 진행할 분석데이터를 선택해 주세요.

3 감성어휘 분석을 진행할 분석데이터를 다운받아 확인해 보세요.

4 분석데이터를 선택한 후, **[적용]** 버튼을 눌러주세요.

※ 감성어휘 분석은 감성분석에 탑재된 형태소 분석기로 어휘 추출을 진행합니다.

키워드정제 페이지에서 직접 편집한 데이터와 차이가 있을 수 있습니다.

문서분류

감성어휘

✓ 적용

다운로드 시 전체를 확인할 수 있습니다.

**5** 텍스트 다운로드 엑셀 다운로드

구분	빈도(건)	감성강도비율(%)	빈도비율(%)
긍정	1301/1570	<b>6</b> 84.33/100.0	82.86/100.0
부정	269/1570	15.66/100.0	17.13/100.0

긍정 키워드	부정 키워드	세부 감정 키워드	긍정 키워드	부정 키워드	세부 감정 키워드
흥미	호감	기쁨			
감정분류	빈도(건)	감정강도	빈도*감정강도	빈도비율(%)	
새롭다	313	2.7778	112.679	19.93	
혁신적	56	3.88889	14.3999	3.56	
기대하다	30	4.66667	6.4285	1.91	
원하다	28	5	5.6	1.78	
특별하다	18	3.77778	4.7647	1.14	
흥미롭다	14	3.4444	4.0645	0.89	
재미있다	9	2.6667	3.3749	0.57	

**5** [텍스트 다운로드], [엑셀 다운로드] 를 클릭하면, 선택한 형식으로 분석 결과 파일이 다운로드 됩니다.

**6** 감성어휘 분석결과로 추출된 감성어휘의 요약 결과를 제공합니다.

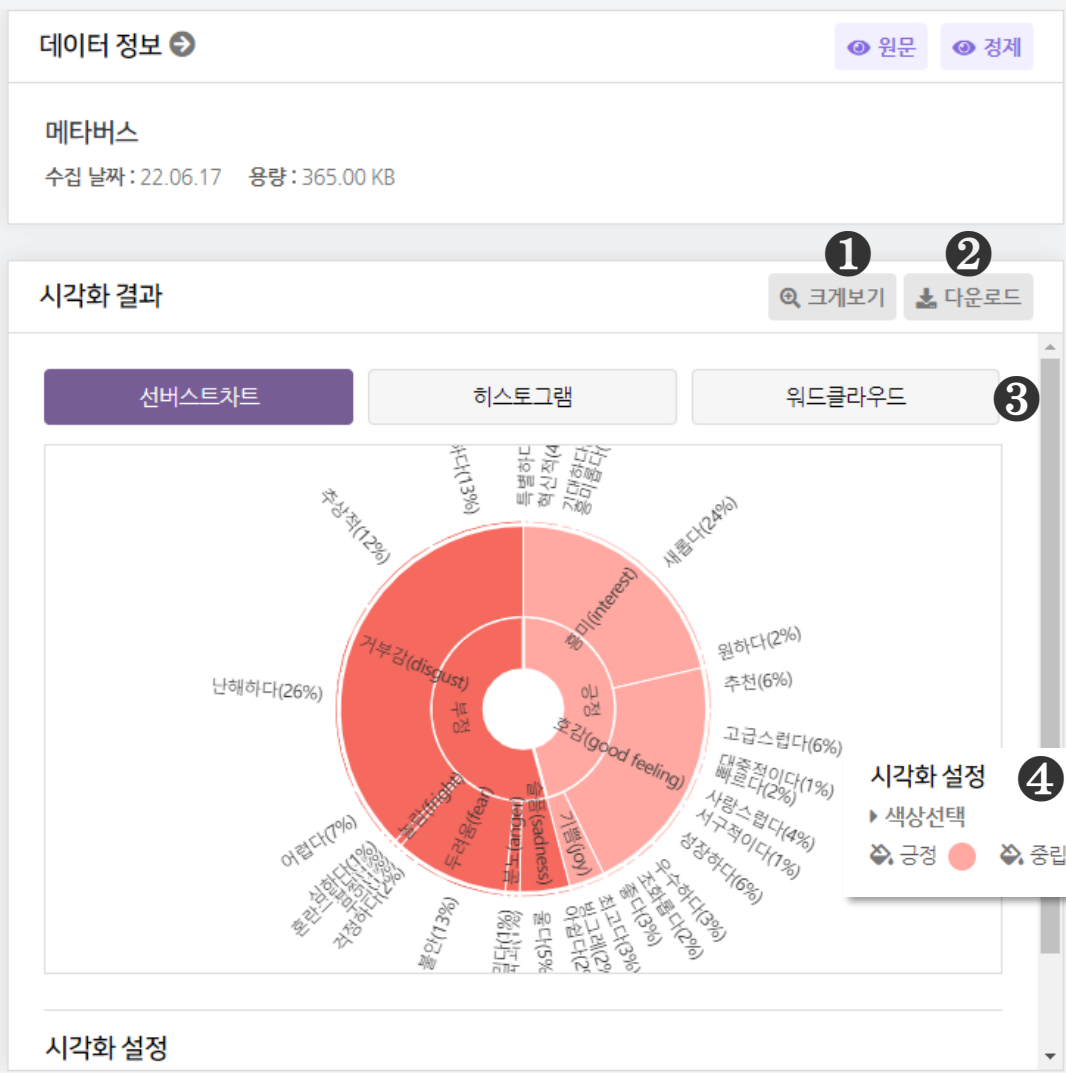
**7** 감성어휘 분석결과로 추출된 감성어휘의 3가지 상세 결과를 제공합니다.

※ 긍정 키워드 : 세부 감정(흥미, 호감, 기쁨) 키워드의 빈도, 감정강도, 빈도\*감정강도, 빈도비율

※ 부정 키워드 : 세부 감정(통증, 슬픔, 분노, 두려움, 놀람, 거부감) 키워드의 빈도, 감정강도, 빈도\*감정강도, 빈도비율

※ 세부 감정 키워드 : 7개 세부 감성분류에 대한 결과로 세부감성 빈도, 세부감성비율

※ 감정강도는 세부 감정(흥미, 호감, 기쁨, 통증, 슬픔, 분노, 두려움, 놀람, 거부감) 안에서 표현의 세기를 의미하며, 7점 Likert 척도로 측정되었습니다.



① [크게보기] 버튼 클릭 시, 시각화가 팝업으로 크게 나타납니다.

② [다운로드] 버튼 클릭 시, 시각화가 다운로드 됩니다.

③ 감성어휘 분석은 3종의 시각화 결과를 제공합니다.

- 선버스트차트, 히스토그램, 워드클라우드

④ [시각화 설정] 을 통해, 시각화의 색상을 변경할 수 있습니다.

**가설검정**

QAP 상관분석    QAP 회귀분석

QAP 상관분석은 상관계수를 바탕으로 두 네트워크 간의 관계를 검증할 수 있습니다. 본 분석에서는 동일한 단어를 구성하고 있는 2개 이상의 네트워크 매트릭스에 대한 상관성을 분석합니다. 예를 들어, '인공지능' 데이터와 '메타버스' 데이터에 대해 동일한 단어로 구성된 네트워크의 상관관계를 분석합니다.

**분석 결과 : QAP Parameters**  
분석 내역이 없습니다. 분석을 진행해주세요.

**분석 결과 : QAP Correlations**  
분석 내역이 없습니다. 분석을 진행해주세요.

**분석 결과 : QAP P-Values**  
분석 내역이 없습니다. 분석을 진행해주세요.

**매트릭스 선택/업로드**  
매트릭스를 선택하거나 업로드를 통해 상관관계를 분석할 매트릭스를 불러올 수 있습니다. 상관관계를 확인할 매트릭스를 2개 이상 선택/업로드해주세요. 단, 모든 매트릭스(1-mode)는 동일한 단어로 구성되어야 합니다.

**파라미터 선택**  
매트릭스 내 노드들 간의 상관성을 측정하기 위한 방법으로 0과 1로 나타내는 이분형 지표이기 때문에 유사성과 상이성 지표로 측정해야 합니다.

유사성 지표  
 피어슨 상관계수     자카드 계수     유클리디안 거리

Permutation

**적용**

❶ [매트릭스 선택/업로드] 영역에서 QAP 상관분석을 진행할 매트릭스를 선택하거나 업로드 합니다.

- 상관관계를 확인하기 위해 매트릭스를 2개 이상 선택/업로드 해주세요.
- 버튼 클릭 시, 파일 업로드를 할 수 있습니다.
- 버튼 클릭 시, 분석 매트릭스를 추가할 수 있습니다.

❷ [파라미터 선택] 에서 [유사성 지표] 와 [Permutation] 을 설정합니다.

- [유사성 지표] : 피어슨 상관계수, 자카드 계수, 유클리디안 거리 중 1개를 선택합니다.
- [Permutation] : 수행할 재배열 횟수를 입력해주세요.

❸ [적용] 버튼을 눌러, 분석을 진행해주세요.

QAP 상관분석

QAP 회귀분석

Measures	Value
1st Data Matrix	야구_매트릭스_40.xlsx
2nd Data Matrix	축구_매트릭스_40.xlsx
3rd Data Matrix	
Matrix Words(N)	39
Permutation Count	3000
Similarity	pearson

### 분석 결과 : QAP Correlations

[텍스트 다운로드](#) [엑셀 다운로드](#)

		1	2
		야구_매트릭스_40.xlsx	축구_매트릭스_40.xlsx
1	야구_매트릭스_40.xlsx	0	0.5993212389729055
2	축구_매트릭스_40.xlsx	0.5993212389729055	0

### 분석 결과 : QAP P-Values

[텍스트 다운로드](#) [엑셀 다운로드](#)

		1	2
		야구_매트릭스_40.xlsx	축구_매트릭스_40.xlsx
1	야구_매트릭스_40.xlsx	0	1.2751816025355835e-43
2	축구_매트릭스_40.xlsx	1.2751816025355835e-43	0

## ④ QAP 상관분석은 3종의 분석 결과를 제공합니다.

### - QAP Parameters

1st Data Matrix	1번째 매트릭스	Matrix Words (N)	분석단어 개수
2nd Data Matrix	2번째 매트릭스	Permutation Count	재배열 횟수
3rd Data Matrix	3번째 매트릭스	Similarity	유사성 지표

### - QAP Correlations : QAP 상관계수

### - QAP P-Values : QAP 상관분석 유의수준



QAP 상관분석

QAP 회귀분석

QAP 회귀분석은 네트워크 매트릭스로 표현된 종속변수에 대한 독립변수 매트릭스의 영향을 평가하는데 사용되는 다중회귀분석(MR-QAP, Multiple Regression-QAP)입니다. 본 분석에서는 동일한 단어를 구성하고 있는 2개 이상의 네트워크 매트릭스(1-mode)에 대한 인과관계를 분석합니다.

분석 결과 : QAP Parameters

분석 내역이 없습니다. 분석을 진행해주세요.

분석 결과 : QAP Regression Coefficients

분석 내역이 없습니다. 분석을 진행해주세요.

매트릭스 선택/업로드

생성해놓은 매트릭스를 선택하거나 매트릭스 파일(xlsx) 업로드를 통해 인과관계를 분석할 매트릭스를 불러올 수 있습니다. 종속변수 매트릭스는 1개, 독립변수 매트릭스는 1개 이상의 복수로 선택/업로드 할 수 있습니다. 단, 모든 매트릭스(1-mode)는 동일한 단어로 구성되어야 합니다.

▶ 종속변수 매트릭스

1

▶ 독립변수 매트릭스

파라미터 선택

▶ Permutation



2

3000

3

적용

① [매트릭스 선택/업로드] 영역에서 QAP 회귀분석을 진행할 종속변수와 독립변수 매트릭스를 선택하거나 업로드 합니다.

- 회귀분석을 위한 종속 및 독립변수 매트릭스를 1개 이상 선택/업로드 해주세요.
-  버튼 클릭 시, 파일 업로드를 할 수 있습니다.
-  버튼 클릭 시, 분석 매트릭스를 추가할 수 있습니다.

② [파라미터 선택] 에서 [Permutation] 을 설정합니다.

- [Permutation] : 수행할 재배열 횟수를 입력해주세요.

③ [적용] 버튼을 눌러, 분석을 진행해주세요.

QAP 상관분석

QAP 회귀분석

QAP 회귀분석은 네트워크 매트릭스로 표현된 종속변수에 대한 독립변수 매트릭스의 영향을 평가하는데 사용되는 다중회귀분석(MR-QAP, Multiple Regression-QAP)입니다. 본 분석에서는 동일한 단어를 구성하고 있는 2개 이상의 네트워크 매트릭스(1-mode)에 대한 인과관계를 분석합니다.

### 분석 결과 : QAP Parameters

[텍스트 다운로드](#) [엑셀 다운로드](#)

Measures	Value
Dependent Data Matrix	data_2023_3_24_101645_5827_.xlsx
1st Independent Data Matrix	data_2023_3_24_101645_9894_.xlsx
1st Independent Data Matrix	data_2023_3_24_101645_4223_.xlsx
Matrix Words (N)	42
Permutation Count	2000
Model R-Square	0.95358
Model P-Value	17656.055786

### 분석 결과 : QAP Regression Coefficients

[텍스트 다운로드](#) [엑셀 다운로드](#)

Variables		Dependent : data_2023_3_24_101645_5827_.xlsx			
Independent		Unstandardized Coefficient	Standardized Coefficient	P-Value	Standard Error
1	data_2023_3_24_101645_9894_.xlsx	1.474962	0.00304	0	0.029587
1	data_2023_3_24_101645_9894_.xlsx	0.389803	0.001783	0	0.018287
	Intercept	-6.581385	0	0	0

④ QAP 회귀분석은 2종의 분석 결과를 제공합니다.

- QAP Parameters

Dependent Data Matrix	종속변수 매트릭스	Matrix Words (N)	분석단어 개수
1st Independent Data Matrix	1번째 독립변수 매트릭스	Permutation Count	재배열 횟수
2nd Independent Data Matrix	2번째 독립변수 매트릭스	Model R-Square	QAP 회귀식의 R-Square
3rd Independent Data Matrix	3번째 독립변수 매트릭스	Model P-Value	QAP 회귀식의 유의수준

- QAP Regression Coefficients

Unstandardized Coefficient	비표준화 회귀계수	P-Value	유의수준
Standardized Coefficient	표준화 회귀계수	Standard Error	표준오차
Intercept	상수항		

The screenshot displays the 'Customization' interface for a word cloud. The top navigation bar includes '워드클라우드' (Word Cloud), '바차트' (Bar Chart), '에고네트워크' (Ego Network), '파이차트' (Pie Chart), '라인차트' (Line Chart), 'N-gram 네트워크' (N-gram Network), '1-way 워드트리' (1-way Word Tree), '트리맵' (Treemap), '매트릭스' (Matrix), and '네트워크' (Network). The main area shows a word cloud with terms like '서비스', '혁신', '공유', '택시', '정부', '문체', '부총리', '산업', '개인', '모빌리티', '기사', '프리미엄', '이용', '정치', '사회', '갈등', '사업', '해결', '생각', '이슈', '상위', '50개', '내림차순', '오름차순', '1-50', '51-100', '90', '47', '32'. The interface is annotated with numbered steps: 1. Selecting 'N-gram 네트워크', 2. Uploading a file and clicking '적용' (Apply), 3. Clicking '다운로드' (Download), and 4. Configuring settings like '모양 선택' (Shape Selection), '직접 입력' (Manual Input), '기본' (Default), '글꼴 선택' (Font Selection), '크기 선택' (Size Selection), '색상 선택' (Color Selection), and '키워드 선정' (Keyword Selection).

① [시각화 유형] 영역에서, 커스터마이징하고 싶은 시각화 유형을 선택합니다.

- 워드클라우드, 바차트, 에고네트워크, 파이차트, 라인차트, N-gram 네트워크, 1-way 워드트리, 트리맵, 매트릭스, 네트워크

② [시각화 파일 업로드] 구간에서는 원하는 모양의 시각화를 만들 수 있습니다.

- 유형에 적합한 데이터 파일을 업로드 하고, [적용] 버튼을 클릭합니다.
- [예시파일 다운로드] 를 활용해 데이터 작성방법을 확인해주세요.

③ 시각화 결과를 확인하고, [다운로드] 버튼을 눌러 파일로 저장할 수 있습니다.

④ 시각화 설정을 통해 시각화 요소의 색상, 크기 등을 변경할 수 있습니다.