

빅데이터 분석 플랫폼 TEXTOM

텍스톰

매뉴얼

TEXTOM

BIG DATA ANALYSIS PLATFORM TEXTOM

CONTENTS

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

INTRO. 목차

→ 바로가기

회원가입 및 로그인

→ 바로가기

대시보드

→ 바로가기

수집

→ 바로가기

정제

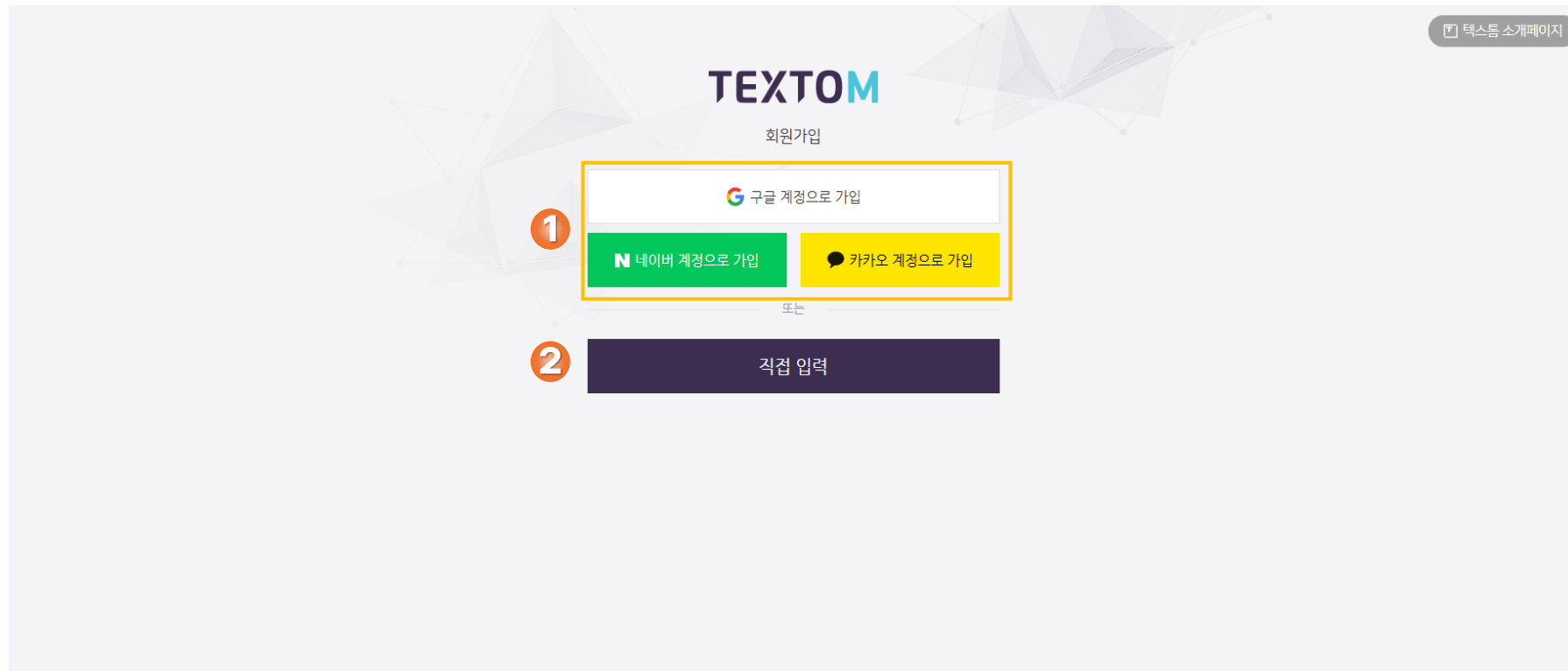
→ 바로가기

분석

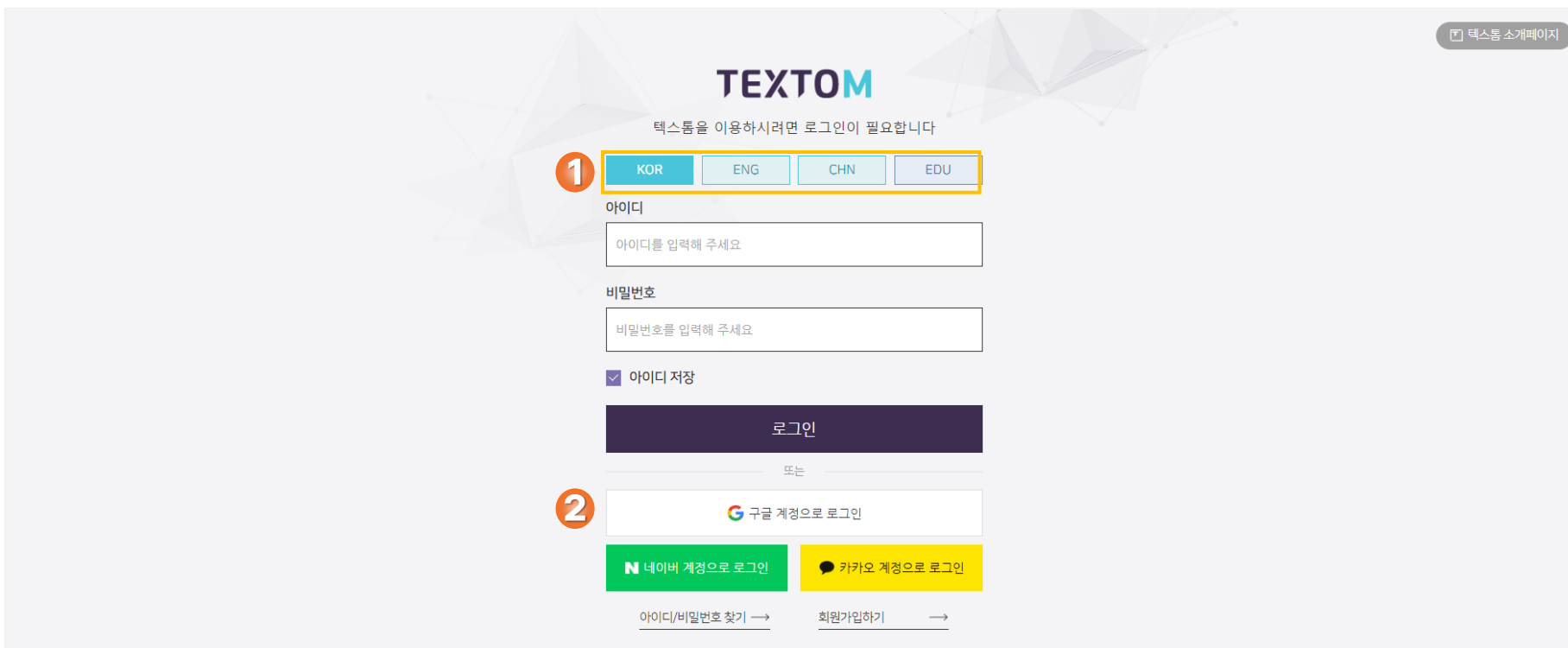
→ 바로가기

기타

→ 바로가기

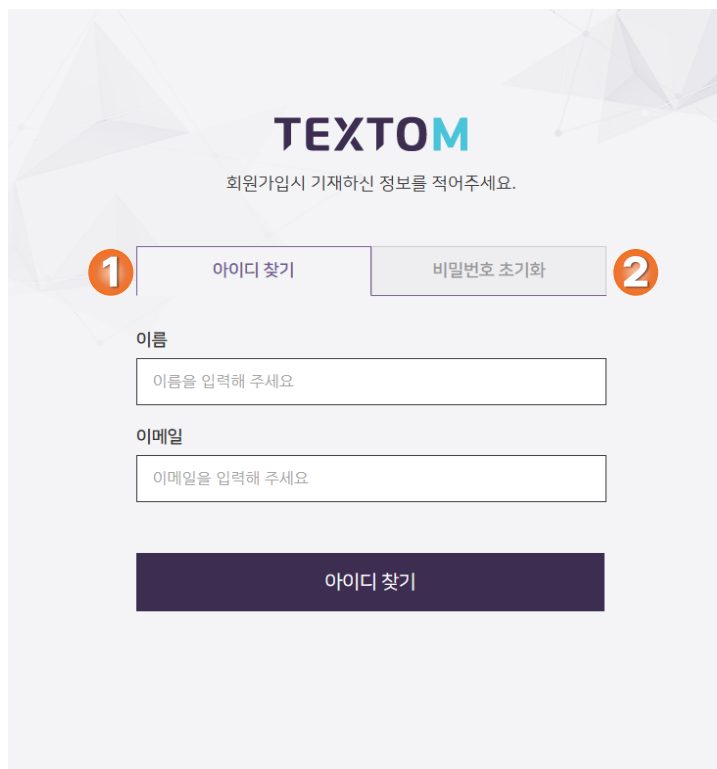


- ① **소셜 계정 회원가입**
기존에 가입되어 있던 구글, 네이버, 카카오계정으로 연동가입이 가능합니다.
- ② **직접 입력 회원가입**
소셜 계정이 없다면 아이디, 비밀번호, 이메일 등 개인정보를 직접 입력하여 회원가입이 가능합니다.



- ① **서비스 선택**
 텍스툼, 텍스툼 Global, 텍스툼 China, 텍스툼 EDU 중 원하는 서비스에 로그인합니다.
 Tip. 기존 회원은 이메일 인증 시, 한·영·중 서비스 동시 이용이 가능합니다.
 Tip. 텍스툼 EDU에 로그인하기 위해선 텍스툼의 회원가입이 먼저 이루어진 후 이용이 가능합니다.

- ② **소셜 계정 로그인**
 구글, 네이버, 카카오 계정으로 로그인합니다.



TEXTOM
회원가입시 기재하신 정보를 적어주세요.

1 아이디 찾기 비밀번호 초기화 2

이름
이름을 입력해 주세요

이메일
이메일을 입력해 주세요

아이디 찾기



TEXTOM

아이디/비밀번호 초기화

비밀번호를 아이디와 동일하게 초기화 하였습니다.
초기화된 비밀번호는 [redacted] 입니다.

3 로그인 페이지로

- ① **아이디 찾기**
이름과 이메일을 통해 텍스트롬 아이디를 찾을 수 있습니다.
- ② **비밀번호 초기화**
아이디, 이름, 이메일을 입력하여 기존 비밀번호를 초기화 할 수 있습니다.
- ③ **비밀번호 재설정**
초기화된 비밀번호는 [마이페이지]-회원정보수정에서 비밀번호를 재설정 할 수 있습니다.
Tip. 초기화된 비밀번호는 이용중이신 텍스트롬 아이디와 동일합니다.

TEXTOM 대시보드

빅데이터팀님 KOR 용량 추가

1 수집·정제 현황 최근1개월 기준 (진행 완료 / 진행 중) **4 빅데이터팀님** 용량 자세히 보기 >

3 / 3 포털 데이터 1 / 1 학술 데이터 3 / 3 보유 데이터 28 / 29 정제 리스트

2 수집리스트 등록된 최근 5개의 수집리스트만 보입니다. 더보기 >

데이터명	수집유형	수집상태	수집일시
data_1.xlsx	보유데이터	수집완료	24.05.28
data_2.xlsx	보유데이터	수집완료	24.05.28
...	보유데이터	수집완료	24.05.28
전세사기	포털/SNS	수집완료	24.05.14
드라마	학술	수집완료	24.05.08

3 정제리스트 등록된 최근 5개의 정제리스트만 보입니다. 더보기 >

데이터명	버전	변경단어	정제상태	정제일시
위성정당.txt	1	0개	진행중	24.05.28
data_2.xlsx	1	0개	진행완료	24.05.28
data_1.xlsx	1	0개	진행완료	24.05.28
!	1	0개	진행완료	24.05.28
전세사기	1	0개	진행완료	24.05.14

잔여용량 757.65 MB / 3.52 GB

5 공지사항 더보기 >

- 공지 [포털/SNS] 수집 방법 변경 안내 (24.05.20 ~) 24.05.20
- 공지 ★ TEXTOM 24 오픈 이벤트 | 숨은 버그를 찾아라 ★ 24.05.02
- 공지 ★ TEXTOM 24 OPEN ★ 24.04.30
- 공지 ★업그레이드를 위한 운영 중단 안내 (일정 변경 - 2024년... 24.03.12
- 공지 ★ TEXTOM GLOBAL 출시★ 24.01.31

텍스톰 매뉴얼 기본 사용법 매뉴얼 활용논문 텍스톰 활용법 및 결과물

FAQ 자주 찾는 질문 Q&A 불편 및 개선사항 의견

6 채팅문의

① 수집 및 정제 현황
포털, 학술, 보유데이터 등 채널별 수집 및 정제 현황을 확인할 수 있습니다.
Tip. 아이콘을 클릭하면 해당 페이지로 이동합니다.

② 수집리스트
최근 등록된 5개 데이터의 수집 진행 현황을 확인할 수 있습니다.
Tip. [더보기] 버튼 클릭 시 해당 페이지로 이동합니다.

③ 정제리스트
최근 등록된 5개 데이터의 정제 진행 현황을 확인할 수 있습니다.
Tip. [더보기] 버튼 클릭 시 해당 페이지로 이동합니다.

④ 잔여 용량 확인
잔여 데이터를 확인할 수 있습니다.
[용량 자세히보기]을 누르면 '마이페이지'로 이동합니다.

⑤ 공지사항, 텍스톰 매뉴얼, 활용논문 등 서비스 정보를 확인할 수 있습니다.

⑥ 채팅문의 시 빠르게 답변 받을 수 있습니다.

데이터 수집

포털/채널 데이터를 수집합니다.

The screenshot shows the TEXTOM data collection interface. On the left is a dark sidebar with navigation options: 대시보드, 수집, 데이터 수집, 수집리스트, 정제, 데이터 전처리, 정제리스트, 분석, 커스터마이징. The main area is titled '데이터 수집' and has tabs for '포털/채널', '학술(RISS)', and '보유데이터'. A '요청채널' button is in the top right. The interface is divided into five numbered steps:

- 키워드 미리보기**: A search bar with '사과' and a '확인' button. A tooltip says '로그인 전 키워드 미리보기를 체험해 보세요.' Below the bar, it says '입력한 키워드의 최근 기간 네이버 검색량을 조회합니다.'
- 수집키워드**: A text input field with a '+' button and a '연산자' dropdown. Below it, it says '키워드초소를 사용하면 여러개의 수집리스트를 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)'
- 기간**: Two date pickers showing '2024-05-07' and '2024-05-14'. Below them, it says '수집기간은 최대 3년까지 설정 가능합니다.'
- 채널**: Three sections for channel selection:
 - NAVER**: Includes checkboxes for 네이버 전체, 블로그, 뉴스, 카페, 지식IN, 웹문서.
 - Daum**: Includes checkboxes for 다음 전체, 티스토리, 뉴스, 카페, 웹문서. A note says '티스토리, 카페, 웹문서는 Daum 정책에 따라 최대 165건까지 수집이 가능합니다.'
 - Google**: Includes checkboxes for 구글 전체, 뉴스, 구글페이스북, 웹문서. A note says '구글페이스북은 구글에서 수집하는 페이스북 문서입니다.'
- 수집리스트생성**: A large blue button with a right-pointing arrow.

On the right side of the interface, there are three charts showing search volume trends for the keyword '사과' over time, with data points for PC and Mobile.

- 키워드 미리보기**
입력한 키워드의 네이버 검색량을 조회합니다.
Tip. 검색량을 기반으로 해당 키워드의 이슈성을 추정할 수 있습니다. 연산자 사용을 지원하지 않습니다.
- 수집키워드**
수집을 원하는 키워드를 입력합니다. [+],[-] 버튼으로 키워드셋을 생성하고, 수집 품질 향상을 위한 연산자 검색을 지원합니다.
Tip. “경제”, “관광”와 같은 포괄적인 수집 키워드 보다는 “경기도 경제”, “안동 관광” 등 세부적인 키워드가 데이터 품질을 향상하고 사용 용량을 절감해 줍니다.
- 수집단위 · 기간**
원하는 수집단위와 기간을 입력합니다. 대용량 데이터 수집 시 ‘요청채널’을 이용해주세요.
Tip. 연관성이 높은 데이터만 간편하게 수집하고 싶어요 -> ‘일’단위 선택
대용량 전체 데이터가 필요해요 -> ‘일’ 단위 선택
- 채널**
수집 포털/채널을 선택합니다.
- 선택한 옵션에 따라 수집을 실행합니다.

TEXTOM

☰ 데이터 수집

백데이터팀님 KOR 용량 추가

포털/채널 학술(RISS) 보유데이터 요청채널

1 수집키워드 수집할 키워드를 입력하세요. +

2 기간 [] - []

3 수집리스트생성 >>

☑ 채팅문의

소개페이지 바로가기

- ① **수집키워드**
수집을 원하는 키워드를 입력합니다.
[+],[−] 버튼으로 원하는 키워드셋을 생성하여 다음의 항목을 수집합니다.

- 제목
- 국문 초록
- 다국어 초록

Tip. “경제”, “관광”와 같은 포괄적인 수집 키워드 보다는 “경기도 경제”, “안동 관광” 등 세부적인 키워드가 데이터 품질을 향상하고 사용 용량을 절감에 좋습니다.

- ② **수집 기간**
학술 데이터 수집은 연 단위로 가능합니다.

- ③ 선택한 옵션에 따라 수집을 실행합니다.

데이터 수집

사용자의 보유 데이터를 업로드합니다.

- <9> -

TEXTOM 데이터 수집

빅데이터팀 KOR 용량 추가

포털/채널 학습(RISS) 보유데이터 요청채널

1 데이터업로드

2

txt,xlsx,csv 파일만 업로드 가능합니다.
한 번에 최대 5개 파일까지 선택이 가능합니다.
※ 분리된 수집리스트로 등록

스프레드시트 파일의 경우 지정된 컬럼에 맞는 데이터를 업로드하여 시계열 분석 등 추가적인 분석 기능을 활용할 수 있습니다.
※ A컬럼 - 제목, B컬럼 - 본문, C컬럼 - 채널, D컬럼 - 날짜
※ 업로드 파일 예시

	제목 컬럼	본문 컬럼	채널 컬럼	날짜 컬럼
	A	B	C	D
1	제목	본문	채널	날짜
2	건설사가 폐배터리를? ...	업계에 따르면 글로벌...	news	2022-01-31
3	연매출 20조 시대 연 ...	또 급증하는 수요에...	news	2022-01-31

위의 양식에 맞지 않는 스프레드 시트 데이터의 경우 분석 대상 데이터는 A컬럼에 입력해주세요.

2 수집리스트생성 >>

소개페이지 바로가기

채팅문의

- ① 데이터 업로드
사용자가 보유한 데이터를 업로드합니다. [+] 버튼을 눌러 최대 5개 파일을 선택할 수 있습니다.

Tip. 엑셀파일을 업로드한 경우, 예시 파일의 형식에 따르면 시계열 분석 등 추가적인 분석 기능을 활용할 수 있습니다. 예시 파일 양식과 맞지 않는 경우, 분석 대상 데이터는 A컬럼에 입력해주세요.


- ② 예시파일
해당 버튼을 클릭하면 예시 형식을 다운로드할 수 있습니다. 참고하여 지정된 형식에 맞게 데이터를 업로드하세요.

- ③ 수집리스트를 생성합니다.


TEXTOM
데이터 수집
지금 회원가입시 10MB 무료 증정
KOR
로그인

포털/채널
학술(RISS)
보유데이터
요청채널


특별한 수집을 원하는 채널과 항목을 자유롭게 요청해보세요!
요청하신 항목을 검토하여 수집 비용과 소요 기간을 신속하게 안내해드립니다.




SNS(트위터)




유튜브



네이버 뉴스/댓글



스마트스토어 리뷰



카페, 디시인사이드 게시판

채널 설명

채널 URL

수집 설명

기간 -

요청 채널 의뢰 >>

소개페이지 바로가기
채팅문의

수집 의뢰

텍스톰이 제공하는 채널 외의 데이터 수집을 원하면 요청채널로 등록해주세요.

- 작성해주신 내용을 바탕으로 요청 사이트의 데이터 수집가능여부 파악 후, 산출법에 따라 견적이 책정됩니다.
- 요청 채널을 통한 수집 방법은 수집된 데이터만 전달 받거나, 요청채널 페이지 내에 구축하는 방법이 있습니다.
- 수집된 데이터는 엑셀(.xlsx) 파일 또는 txt 파일로 전달됩니다.
- 구축이 완료되면 수집하기에서 제공하는 기존 채널과 동일한 방법으로 이용하시면 됩니다.

수집리스트

포털/채널에서 수집한 데이터 셋을 확인할 수 있습니다.

The screenshot shows the '수집리스트' (Collection List) page in the TEXTOM system. On the left is a navigation menu with options like '대시보드', '수집', '정제', '분석', and '커스터마이징'. The main area displays a table of collected data sets with columns for '키워드' (Keyword), '채널' (Channel), '기간' (Period), '수집날짜' (Collection Date), '용량' (Volume), and '수집상태' (Collection Status). A search bar at the top left is labeled '1 키워드검색'. The first row is selected, and a detailed view on the right shows '2 전세사기' (Real Estate Scams) with a '3 삭제' (Delete) button. This view includes a '데이터 선택' (Data Selection) section with a date range '2024-03-01 ~ 2024-05-14' and a table of selected items with columns for '채널' (Channel), '섹션' (Section), '수집량(건)' (Collection Count), '용량' (Volume), and '원문' (Original Document). A '4 미리보기' (Preview) button is visible. Below this is a '5 데이터 전처리하기' (Pre-process Data) button with a note '(예상 차감 용량: 1.53 MB)'. At the bottom, a '6 잔여 용량: 757.65 MB' (Remaining Capacity: 757.65 MB) is shown with a '용량 추가' (Add Capacity) button.

① 키워드를 통해 검색이 가능합니다.

② **수집데이터 조회**
수집데이터 목록을 보여줍니다. 체크박스를 클릭하면 우측 영역에 해당 데이터의 정보를 조회할 수 있습니다.

③ **데이터 삭제**
체크박스 선택 후 [삭제] 버튼을 누르면 수집리스트에서 삭제됩니다.
Tip. 수집 완료 이후 30일이 경과하면 자동 삭제됩니다. 필요한 데이터는 다운로드하는 것을 권장합니다.

④ **데이터 정보 확인**
데이터를 클릭하면 미리보기 영역에 해당 데이터의 채널, 섹션, 수집량, 원문 미리보기 등의 정보를 제공합니다.
Tip. 원문을 미리 확인하면 데이터 전처리 단계에서 스팸 키워드 필터링에 도움이 됩니다.

⑤ 해당 데이터의 전처리를 진행합니다.

⑥ 예상 차감 용량과 잔여 용량을 확인하고 용량을 추가할 수 있습니다.

TEXTOM 수집리스트

포털 | 학술(RISS) | 보유데이터

1 키워드검색

수집이 완료된 데이터는 30일이 지나면 삭제됩니다.

1개 선택 / 총 1 개

2

<input type="checkbox"/>	키워드	채널	기간	수집날짜	용량	수집상태
<input checked="" type="checkbox"/>	드라마	RISS	2024~2024	2024-05-08	531.70 KB	수집완료

3 삭제

4 데이터 선택

▶ 드라마
2024 ~ 2024

채널	수집량(건)	용량	원문
RISS	166	531.70 KB	미리보기

5 데이터 전처리하기»

(예상 차감 용량: 531.70 KB)

간여 용량을 확인해주세요!

6 간여 용량: 757.65 MB

용량 추가

☑ 채팅문의

① 키워드를 통해 검색이 가능합니다.

② **수집데이터 조회**
수집데이터 목록을 보여줍니다.
체크박스를 클릭하면 우측 영역에 해당
데이터의 정보를 조회할 수 있습니다.

③ **데이터 삭제**
체크박스 선택 후 [삭제] 버튼을 누르면
수집리스트에서 삭제됩니다.
Tip. 수집 완료 이후 30일이 경과하면 자동
삭제됩니다.

④ **데이터 정보 확인**
데이터를 클릭하면 미리보기 영역에
해당 데이터의 채널, 섹션, 수집량, 원문
미리보기 등의 정보를 제공합니다.
Tip. 원문을 미리 확인한 후, 전처리
단계에서 필터링 기능을 사용해 원하지
않는 키워드나 데이터를 제거하세요.

⑤ 해당 데이터의 전처리를 진행합니다.

⑥ 예상 차감 용량과 잔여 용량을 확인하고
용량을 추가할 수 있습니다.

수집리스트

보유데이터 수집 결과를 확인하세요.

The screenshot displays the '수집리스트' (Collection List) page in the TEXTOM system. The interface is divided into several sections:

- Header:** Includes user information (백데이터팀님), language (KOR), and a '용량 추가' (Add Capacity) button.
- Navigation:** A sidebar on the left contains menu items: 대시보드, 수집 (selected), 데이터 수집, 수집리스트, 정제, 데이터 전처리, 정제리스트, 분석, and 커스터마이징.
- Filtering:** A search bar (키워드검색) and a '삭제' (Delete) button are present. A note indicates that data older than 30 days will be deleted.
- Table:** A table lists collected files with columns for '키워드' (Keyword), '수집날짜' (Collection Date), '용량' (Capacity), and '수집상태' (Collection Status). The file '위성정당.txt' is selected.
- File Selection Panel:** A '데이터 선택' (Data Selection) panel shows details for the selected file, including a '미리보기' (Preview) button and a '데이터 전처리하기' (Pre-process Data) button.
- Summary:** A summary box shows '잔여 용량: 757.65 MB' (Remaining Capacity: 757.65 MB) and another '용량 추가' button.

- ① 키워드를 검색하여 원하는 데이터를 찾을 수 있습니다.
- ② **수집데이터 조회**
수집데이터 목록을 보여줍니다. 체크박스를 클릭하면 우측 영역에 해당 데이터의 정보를 조회할 수 있습니다.
- ③ **데이터 삭제**
체크박스 선택 후 [삭제] 버튼을 누르면 수집리스트에서 삭제됩니다.
Tip. 수집 완료 이후 30일이 경과하면 자동 삭제됩니다.
- ④ **데이터 정보 확인**
데이터를 클릭하면 미리보기 영역에 해당 데이터의 채널, 섹션, 수집량, 원문 미리보기 등의 정보를 제공합니다.
Tip. [미리보기]를 통해 원문을 미리 확인하면 데이터 전처리 단계에서 스팸 키워드 필터링에 유용합니다.
- ⑤ 해당 데이터의 전처리를 진행합니다.
- ⑥ 예상 차감 용량과 잔여 용량을 확인하고 용량을 추가할 수 있습니다.

- ① **분석 컬럼 선택**
 분석을 진행할 컬럼을 선택합니다.
 Tip. '제목'만 분석하는 경우는 빠른 트렌드 파악과 주제 분류에 적합합니다. 심층적인 분석을 위해서는 전체 텍스트를 분석하는 것이 좋습니다.
- ② **형태소분석**
 이미 정제한 데이터를 업로드 하거나 원문을 그대로 분석하려는 경우에는 [선택안함]으로 설정하세요.
- ③ **분석 언어를 선택합니다.**
- ④ **텍스툼은 Mecab을 커스터마이징한 형태소분석기인 'Mecab-IMC'를 사용합니다.**
 Tip. Mecab-IMC는 띄어쓰기와 상관 없이 사전을 참조하여 어휘를 구분합니다.
- ⑤ **분석 품사를 선택합니다.**
 Tip. 동사 : 주체 간 상호작용, 행위 분석
 형용사 : 감성분석
 외국어 숫자 : 제품명, 고유명사 구분
- ⑥ **사용자 사전**
 특정 주제나 키워드를 여러 번 분석할 경우, 사용자 사전을 적용하여 반복적인 작업 없이 빠르고 효율적인 분석이 가능합니다.
 Tip. 정제할 키워드를 사용자사전에 먼저 등록해주세요. (마이페이지-사용자사전)
- ⑦ **[적용] 버튼을 눌러 선택한 옵션을 전처리 이력에 저장합니다.**

1 데이터 정제 옵션선택 선택안함

2 중복제거 원전일치 선택안함
컬럼 선택 전체 제목 본문

3 스팸 필터링 키워드 선택안함
컬럼 선택 전체 제목 본문
키워드 포함 필터링할 키워드를 입력하세요

4 적용 >>

5 전처리 이력
전제시기 2024-03-01 ~ 2024-05-14 미리보기

6 전처리 이력 삭제

가능명	상세내역
<input type="checkbox"/> 컬럼선택	title, contents
<input type="checkbox"/> 분석언어	ko
<input type="checkbox"/> 분석기	mecab
<input type="checkbox"/> 분석품사	noun
<input type="checkbox"/> 사용자 사전	선택안함

7 데이터 전처리하기 >>

소개페이지 바로가기

채팅문의

선택한 컬럼을 기준으로 특정 단어를 포함한 데이터를 남기거나, 제거합니다.

1. 키워드 포함 : 해당 단어를 포함한 데이터만 분석에 활용합니다.
2. 키워드 제외 : 해당 단어를 포함한 데이터를 분석에서 제외합니다.
3. AND : 입력한 키워드 포함·제외 조건을 '모두 만족' 하도록 필터링합니다.
4. OR : 입력한 키워드 포함·제외 조건을 '하나만 만족' 하여도 필터링합니다

- ① **데이터 정제 선택**
중복제거, 스팸 필터링 등의 정제 기능을 사용할 수 있습니다.
Tip. 데이터 용량 사용 절감과 효율적인 분석을 위해 사용하는 것을 권장합니다.
- ② **중복제거**
선택한 컬럼을 기반으로 중복 데이터를 제거합니다. 복수 조건 등록 시 각 컬럼에서 동일한 데이터가 삭제됩니다.
- ③ **스팸 필터링**
선택한 컬럼을 기준으로 포함하거나 제외할 키워드를 사용하여 데이터를 필터링합니다.
- ④ [적용] 버튼을 눌러 선택한 옵션을 전처리 이력에 저장합니다.
- ⑤ **데이터 미리보기**
분석할 데이터를 미리 확인할 수 있습니다.
Tip. 원문을 미리 확인하면 스팸 키워드 필터링과 데이터 용량 절감에 도움이 됩니다.
- ⑥ 체크박스를 선택하여 [삭제] 버튼을 클릭하면 적용한 분석 옵션을 삭제할 수 있습니다.
- ⑦ 선택한 분석 옵션으로 데이터 전처리를 진행합니다.

정제리스트

키워드 정제 현황을 확인하고 관리합니다.

- ① [편집] 버튼을 눌러 데이터명을 수정할 수 있습니다.
- ② 키워드 정제 목록 키워드를 편집한 이력을 확인할 수 있습니다.
- ③ 바로 편집 파일 업로드 없이 웹에서 바로 정제를 할 수 있습니다.
- ④ 데이터 다운로드 원문 데이터, 정제 데이터, 날짜 등 관련 정보를 담은 종합 파일을 엑셀(.xlsx) 형식으로 다운로드할 수 있습니다.
- ⑤ 정제 버전 및 이력에 대한 간단한 메모를 남길 수 있습니다.
- ⑥ 정제 파일 업로드 ④의 파일을 다운로드하여 사용자가 직접 정제한 후 업로드하면 새로운 버전을 추가할 수 있습니다.
Tip. 반드시 ④에서 다운로드한 파일 형식을 유지해야 합니다.
- ⑦ 정제 이력 삭제 체크박스로 삭제할 정제 이력을 선택한 후 삭제합니다.
- ⑧ 새로고침 정제 현황을 실시간으로 확인합니다.

정제리스트

정제리스트 및 데이터를 관리합니다.

TEXTOM

≡ 정제리스트

백데이터팀 KOR [공인 추가]

키워드검색

총 10 건

키워드 정제 목록

0개 선택 / 총 1 개

정제 임포트 삭제

키워드	정제일	버전	삭제
data_2.xlsx	24-05-28	1	삭제
data_1.xlsx	24-05-28	1	삭제
트랙스크로스오버.txt	24-05-28	1	삭제
전세사기	24-05-14	1	삭제
드라마	24-05-14	1	삭제
"증형세단"	24-05-08	2	삭제
"단전호흡"	24-05-02	3	삭제
"단전호흡"	24-05-02	1	삭제
홍선	24-05-01	1	삭제
위성정당.txt	24-05-01	1	삭제

키워드 정제 목록 상세

0개 선택 / 총 1 개

키워드	정제일	버전	삭제
전세사기	2024-05-14 11:35:56	13개	미지정

1 전처리 정보

2 복사 다운

수집정보

키워드	전세사기	수집 기간	2024-03-01 ~ 2024-05-14
수집 채널	네이버(블로그,뉴스)	수집 범위	제목, 본문, URL
수집 단위	월 단위	수집 일시	2024-05-14
수집 용량	1.53 MB		

형태소 분석

분석기	mecab
분석품사	noun,adjective
사용자 사전	-

데이터 정제

중복제거	적용(title, contents)
스팸 필터링	-

3 키워드검색

4 데이터명 정제일 버전

5 데이터 분석하기

채팅문의

- 1 전처리 정보
데이터의 수집 및 정제 정보를 조회할 수 있습니다.
- 2 전처리 정보를 표 형식으로 복사하거나 다운로드 할 수 있습니다.
- 3 키워드 검색
키워드로 정제리스트를 검색할 수 있습니다.
- 4 정제리스트 관리
정제리스트 조회 및 삭제 등 관리 작업을 할 수 있습니다.
- 5 선택한 데이터로 분석을 진행합니다.

키워드 정제

데이터 업로드 없이 웹 페이지에서 키워드 정제를 진행합니다.

키워드 정제

1 원문/정제 데이터

원문	정제
1 전세사기 신고 전세 보증금 미 반환 피해자 높 전 전세사기 피해 발생 해로 법 대응 부동산 투자 과정 형성 방식 선호 전세 계약 시기 전세 보증금 미 반환 금전 피해 발생 전세금 경우	
2 부동산 전세사기 법률 논점 전세사기 종류 우수 사기 사건 임대인 위임 관리자 월세 약정 세입자 속 전세 계약 진척 중간 보증금 일 세 지불	
3 전세사기 예방 교육 음악회 금요일 저녁 전세사기 예방 교육 분 퇴근 저녁 샌 박근 하나 버스 분위 똑같 일 주일 만 강사 재미 있 것 회차 종료 시 간 짧 내용	
4 계 전세사기 피해자 의뢰인 저 집 주인 세 전세 계약 체결 문제 부동산 등기부갑 소용 당사 거기 수사 과정 써 자기 사기죄 처벌 수 것 판단 전세사 기 경우 예방	
5 전세사기 방지 임대인 문제 상황 사기 사건 신속 대응 먼저 전세사기 방지 임대인 구축 실행 위기 전문가 투자 부동산 경기 좋 때 문제 집값 상태 세 입자 때 기존 전세	
6 전세 작업 대출 대출 사기 집행유예 대응 실행 처벌 전세사기 작업 대출 연루 돈 시기 사람 은행 돈 직업 없 처벌 수 것 영두 전세 작업 대출 대출 사 기 실행 처벌 위기 집행유예 등 응방 마련 필요	
7 전세 보증금 미 반환 전세 보증보험 사기 현명 대처 방안 전세사기 대비 보증보험 가입 필요 전세사기 대비 전세 보증보험 사기 대비 나라 인종 보 증보험 가입 추천 해당 보험 대표 것 한국 주택 금융 공사 주택 도시 보증 공사 서울보증보험 등 있	
8 전세사기 피해자 전세 보증금 미 반환 문제 해결 다음 내용 임차 주택 관련 경매 공매 진행 사기 범행 의도 확인 전세사기 피해자 변호인 조력 대응 전세 보증금 반환 소송 진행 입증 것 부분	
9 임차 보증금 반환 소송 부동산 전세사기 신속 대응 부동산 전세사기 강릉 전세 문제 많 임대차 기간 만료 임대인 보증금 반환 분 부동산 전세사기 사기죄 고소 방법 있 임대차 계약 당시 보증금 편취 의사 것	
10 전세사기 보증금 방법 가지 저 민사법 형사법 전문 분야 자격 취득 년 차 변호사 안전 전세사기 문제 고민 중 본 부당 이득 사람 사기죄 형량죄 등 형사상 책임 수 책임 소재 여립	
11 전세사기 변호사 비용 보증금 필독 형사 성공 사례 하는 확인 법원 법원 서울 형사 사건 대표 성공 사례 대어금 사기 고소 의뢰인 대리 무죄 판결 이 후 기 전세사기 변호사 문제 사례 필독 임대인 전세금 개선 기 회보 의향	

2 분석 결과

TF	N-gram
1 전세사기 7436	
2 전세 4085	
3 피해자 1659	
4 보증금 1598	
5 사기 1535	

3 단어 변경

단어 직접입력

AI 자동정제 단어

변경 제거

변경할 단어를 입력하세요.

수정할 단어를 입력하세요.

4 수정내역

사용자 사전

정제 전	정제 후
<input type="checkbox"/> 전세 사기	전세사기
<input type="checkbox"/> 보증 보험	보증보험
<input type="checkbox"/> 공인 중개사	공인중개사
<input type="checkbox"/> 개인 회성	개인회성
<input type="checkbox"/> 갑 투자	갑투자
<input type="checkbox"/> 국토 교통부	국토교통부

5 + 신규버전 생성

6 ✓ 기존버전 적용

- ① 원문/정제 구분 탭
원문데이터와 정제 탭을 전환하여 데이터를 확인할 수 있습니다.
- ② 분석 결과 확인
TF(단어빈도)와 N-gram 분석 결과를 확인하여 단어 정제에 참고할 수 있습니다.
- ③ 단어 변경
단어를 직접 입력하여 변경하거나 AI 자동 정제 기능을 활용하여 간편하게 정제할 수 있습니다.
- ④ 수정내역
단어 수정 내역을 확인하고 삭제할 수 있습니다. 정제하고 있는 수정 내역을 사용자 사전에 등록하거나, 기존의 사용자 사전을 불러올 수 있습니다.
- ⑤ 신규 버전 생성
수정 내역을 신규버전으로 추가할 수 있습니다.
- ⑥ 기존 버전 적용
수정 내역을 기존 정제 버전에 적용할 수 있습니다.
Tip. 변경 내용을 적용한 후에는 수정 전으로 되돌릴 수 없으니, 수정내역 검토 및 정제데이터를 미리 내려 받아 두는 것을 추천합니다.

키워드 정제

원문/정제 데이터와 분석 결과를 참고하여 키워드를 정제합니다.

키워드 정제

1 원문/정제 데이터

원문 | 정제

- 건 검색 / 총 3259 건

2 검색할 단어

1	전세사가 신고 전세 보증금 미 반환 피해자 농건 전세사가 피해 발생 배로법 대응 부동산 투자 과정 형성 방식 선호 전세 계약 시기 전세 보증금 미 반환 골건 피해 일 발생 전세금 경우
2	부동산 전세사가 법률 논점 전세사가 종류 우수 것 시기 사건 임대인 위임 관리자 월세 약정 세입자 측 전세 계약 진척 중간 보증금 일 세 지불
3	전세사가 예방 교육 음악회 금요일 저녁 전세사가 예방 교육 분 퇴근 저녁 선택 퇴근 하나 버스 분 뒤 똑같은 주말 만강사 재미 있 것 회차 종료 시 간 짧 나중
4	계 전세사가 피해자 의뢰인 씨 집 주인 씨 전세 계약 체결 문제 부동산 등기부 집 소유 탁사 거기 수사 과정 씨 자기 사기죄 처벌 수 것 판단 전세사 기 경우 예방
5	전세사가 바지 임대인 문제 상환 시기 사건 신속 대응 먼저 전세사가 바지 임대인 구속 실행 위기 전문가 투자 부동산 경기 좋 때 문제 집값 상태 세 입자 때 기존 전세
6	전세 작업 대출 대출 사기 집행유예 대응 실행 처벌 전세사가 작업 대출 연루 든 시기 사람 은행 든 직업 없 처벌 수 것 연두 전세 작업 대출 대출 사 기 실행 처벌 위기 집행유예 등 응방 마련 필요
7	전세 보증금 미 반환 전세 보증보험 사기 현명 대처 방안 전세사가 대비 보증보험 가입 필요 전세사가 대비 전세 보증보험 사기 대비 나라 인종 보 증보험 가입 추천 해당 보험 대표 것 한국 주택 금융 공사 주택 도시 보증 공사 서울보증보험 등 있
8	전세사가 피해자 전세 보증금 미 반환 문제 해결 다음 내용 입학 주택 관련 경매 공매 진행 시기 범행 의도 확인 전세사가 피해자 변호인 조력 대응 전세 보증금 반환 소송 진행 입증 것 부분
9	임차 보증금 반환 소송 부동산 전세사가 신속 대응 부동산 전세사가 정통 전세 문제 많 임대차 기간 만료 임대인 보증금 반환 부동산 전세사가 사기죄 고소 방법 있 임대차 계약 당시 보증금 편취 의사 것
10	전세사가 보증금 방법 가지 저 민사법 행사법 전문 분야 자격 취득 년 차 변호사 안산 전세사가 문제 고민 응 본 부당 이득 사들 사기죄 형량 징동 형 사상 책임 수 책임 소재 어렵
11	전세사가 변호사 비용 보증금 필독 형사 성공 사례 한눈 확인 법무 법인 서울 형사 사건 대표 성공 사례 대여금 사기 고소 의뢰인 대리 무죄 판결 이 후 기 고소인 변호사 주택 세력 피소 의뢰인 전세금 계산 오차 차분 인정

+ 신규버전 영상 ✓ 기존버전 적용

3 분석 결과

TF | N-gram

- 건 검색 / 총 6294 건

1	전세사가	7436
2	전세	4085
3	피해자	1659
4	보증금	1598
5	사기	1535

3

TF | N-gram

- 건 검색 / 총 53816 건

Bigram

1	전세사가	피해자	1084
2	전세	보증금	740
3	전세사가	피해	677
4	전세사가	변호사	582
5	보증금	반환	490

국토교통부 국토교통부

① 원문/정제 구분 탭
원문데이터와 정제 탭을 전환하여 데이터를 확인할 수 있습니다.

② 키워드 검색
특정 키워드가 원문과 정제 데이터에서 어떻게 활용되고 있는지, N-gram을 통해 특정 키워드가 어떤 단어와 함께 사용되고 있는지 파악하여 정제하는데 활용할 수 있습니다.
Tip. 데이터 정제 과정에서 데이터 확인, 연결된 단어, 수정된 단어가 적용된 결과 등을 확인할 때 활용됩니다.

③ 분석 결과 확인
TF(단어빈도)와 N-gram 분석 결과를 확인할 수 있습니다.
Tip. 단어빈도와 N-gram 분석결과를 참고하면 수정할 유의어, 제거할 불용어, 결합할 단어 찾기에 효과적입니다.

키워드 정제

다양한 기능을 활용하여 키워드를 정제합니다.

The screenshot shows the '키워드 정제' (Keyword Cleaning) interface. It includes a '원문/정제 데이터' (Original/Cleaned Data) table on the left, a '단어 변경' (Word Change) dialog box, a '분석 결과' (Analysis Results) table, and a '수정내역' (Modification History) table. The interface is annotated with four numbered callouts:

- 1**: '단어 변경' dialog box with '단어 직접입력' (Direct Input) and 'AI 자동정제 단어' (AI Auto-Cleaned Words) tabs. It includes a search field for words to change and a '+ ' button to add them.
- 2**: '단어 변경' dialog box with '단어 직접입력' and 'AI 자동정제 단어' tabs. It includes a search field for words to change and a '+ ' button to add them.
- 3**: '단어 변경' dialog box with '단어 직접입력' and 'AI 자동정제 단어' tabs. It includes a table showing the results of the cleaning process.
- 4**: '수정내역' (Modification History) table showing the changes made to the keywords.

- ① **단어 직접 변경**
 변경할 단어와 수정할 단어를 직접 입력하여 변경할 수 있습니다. 여러 단어를 하나의 수정 단어로 변경하고자 할 때, [+] 버튼을 클릭하여 변경할 단어를 추가할 수 있습니다.
- ② **키워드 및 데이터 제거**
 특정 단어를 전체 데이터에서 제거할 수 있습니다. [+] 버튼을 클릭하여 제거할 단어를 추가할 수 있습니다. 특정 키워드가 포함된 데이터를 제거하고자 할 때는 체크박스를 선택하세요.
- ③ **AI 자동 정제**
 SI를 기반으로 단어가 연속적으로 나타날 확률을 계산하여 최대 200개의 단어쌍을 추천합니다.
- ④ 버튼을 눌러 변경한 단어를 수정내역에 적용합니다.

데이터 분석

분석리스트 조회하고 데이터의 세부 정보를 확인합니다.

The screenshot shows the '데이터 분석' (Data Analysis) section of the TEXTOM interface. It includes a sidebar with navigation options like '대시보드', '수집', '정제', '분석', and '커스터마이징'. The main content area is divided into several sections:

- 3 전세사기**: A search bar for '전세사기' (Real Estate Scams).
- 4 분석 정보**: A table showing analysis statistics for '전세사기'.

정제버전	형태소분석기	명사	형용사	동사	외국어	숫자	기타	합계
1	Mecab	6180개	118개	0개	0개	0개	0개	6298개
- 5 분석 내역**: A table listing analysis methods and their status.

구분	세부기능	바로가기
단어분석	단어빈도	분석하기
	N-gram	분석하기
	TF-IDF	분석하기
매트릭스	시계열 분석	분석하기
	매트릭스 1-mode	분석하기
	매트릭스 2-mode	분석하기
네트워크 분석	네트워크 속성	분석하기
	중심성	분석하기
	예고 네트워크	분석하기
	최단경로	분석하기

On the right side, there is a '분석리스트' (Analysis List) table with columns for '데이터명', '정제일', and '버전'. A search bar at the top allows for keyword searches. A '데이터 분석하기' button is located at the bottom of the list.

- 키워드를 검색하여 해당 단어를 포함하는 데이터를 확인할 수 있습니다.
- 분석리스트**
정제가 완료된 데이터를 조회하고 삭제할 수 있습니다. 특정 데이터를 클릭하면 좌측영역에 해당 데이터의 정보가 나타납니다.
Tip. [>] 버튼을 클릭하면, 분석리스트 창을 숨길 수 있습니다.
- 데이터명 수정**
[수정] 버튼을 클릭하여 사용자가 원하는 이름으로 데이터의 명칭을 쉽게 수정할 수 있습니다.
- 분석 정보**
해당 데이터의 형태소 분석 결과를 확인할 수 있습니다.
Tip. 정제버전을 선택할 수 있으며, 분석 정보를 복사하거나 엑셀(.xlsx)파일로 다운로드 할 수 있습니다.
- 분석 내역**
해당 데이터에 대한 분석 결과를 조회할 수 있습니다. [분석하기] 버튼을 클릭하면 개별 분석 결과로 이동합니다.
- 선택한 데이터의 분석 결과 화면으로 이동합니다.

단어분석

단어 빈도(TF) 분석을 진행할 단어를 선택합니다.

단어 빈도(Term Frequency)

전체 문서에서 특정 단어가 언급된 횟수를 계산합니다.
단어 언급빈도는 중요도 분석을 통해 문서의 주제 또는 태도나 감성을 추론하는 지표로 활용될 수 있습니다.

상위 20개 단어 선택

0개 선택 / 총 6294 개

단어	빈도(건)	백분율(%)	누적백분율(%)
전세사기	7436	6.379	6.379
전세	4085	3.504	9.883
피해자	1659	1.423	11.306
보증금	1598	1.371	12.677
사기	1535	1.317	13.994
피해	1400	1.201	15.195
변호사	1178	1.011	16.206

시각화 결과

워드클라우드 버블차트 세로막대그래프 가로막대그래프

원쪽 리스트에서 시각화 할 단어를 선택해주세요.

- 분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- 상위 단어 자동 선택**
전체 단어 중 상위 20개, 50개, 75개, 100개 단어를 자동으로 선택합니다.
- 단어 선택**
최상단의 체크박스를 클릭하면, 상위 100개의 단어를 자동으로 선택합니다.
단어를 직접 선택하려면 해당 단어 옆의 체크박스를 사용하세요.
Tip. 백분율은 전체 언급량을 100으로 했을 때 언급된 양을 의미합니다.
- 분석 결과 다운로드**
[선택 다운로드], [전체 다운로드]를 클릭하면, 선택한 단어 또는 전체 단어의 빈도, 백분율, 누적 백분율을 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.
- 선택한 단어를 시각화 합니다.
Tip. 가시성을 위해 선택 단어 수를 50개~70개의 단어를 선택하는 것을 추천합니다.
- 시각화 결과를 나타냅니다.

단어분석

단어 빈도(TF) 분석 결과를 시각화 합니다.

1 워드클라우드

2 시각화 설정

3 시각화 설정

4 크게보기

5 다운로드

- ① **시각화 결과**
단어빈도 분석은 총 4종의 시각화 결과를 제공합니다.

워드클라우드	버블차트
가로막대차트	세로막대차트

- ② **시각화 설정**
사용자는 시각화 결과의 속성을 원하는 대로 변경할 수 있습니다.
Tip. 워드클라우드 : 모양, 글꼴, 색상 버블차트 : 배경색, 글자색, 글자크기
세로.가로막대그래프 : 바 색상, 범례위치

- ③ **모양 직접 선택 (워드 클라우드)**
워드클라우드 사용 시, 특정 모양으로 시각화를 생성할 수 있습니다.
Tip. 샘플이미지를 참고하여 원하는 모양의 이미지를 업로드하세요.

- ④ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.

- ⑤ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

단어분석

N-gram 분석을 진행할 단어를 선택합니다.

- <24> -

TEXTOM 데이터 분석 콘솔

백데이터팀님 데이터 추가

단어분석 매트릭스 네트워크 분석 토픽 분석 감성분석 QAP 가설검정

단어 빈도 N-gram TF-IDF 시계열

1 분석 알고리즘

문자열에서 N개의 연속된 요소를 추출하는 통계 기반 언어 분석모델로, 문장 내 연속하여 동시 출현하는 단어와 그 빈도를 계산합니다. 문장을 몇 개의 단어 개수에 따라 나눌지에 따라 종류가 결정됩니다.

2 Bigram(2) Trigram(3) Tetragram(4)

3 상위 20개 단어 선택

0개 / 총 53816 개 ※전체 다운로드는 최대 20만 건까지 가능합니다.

4

단어1	단어2	빈도(건)
전세사기	피해자	1084
전세	보증금	740
전세사기	피해	677
전세사기	변호사	582
보증금	반환	490
전세	계약	477

5 선택 다운로드 전체 다운로드

6 선택단어 시각화 >>

7

시각화 결과

원쪽 리스트에서 시각화 할 단어를 선택해주세요.

시각화 설정

색상선택

선택상 화살표 모양 글자 크기 14

- 1 분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- 2 N-gram 범위 설정**
연속으로 나타나는 단어의 개수를 2개(Bigram)에서 4개(Tetragram) 사이에서 선택할 수 있습니다.
- 3 상위 단어 자동 선택**
전체 단어 중 상위 20개, 50개, 75개, 100개 단어를 자동으로 선택합니다.
- 4 N-gram 단어 선택**
최상단의 체크박스를 클릭하면, 상위 100개의 단어를 자동으로 선택합니다. 단어를 직접 선택하려면 해당 단어 옆의 체크박스를 사용하세요.
- 5 분석 결과 다운로드**
[선택 다운로드], [전체 다운로드]를 클릭하면, 선택한 N-gram 또는 전체 N-gram의 빈도를 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.
- 6** 선택한 단어를 시각화 합니다.
- 7** 시각화 결과를 나타냅니다.

단어분석

N-gram 분석 결과를 시각화 합니다.

TEXTOM
데이터 분석 콘솔
백데이터팀님
데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

단어 빈도
N-gram
TF-IDF
시계열

N-gram 분석 알고리즘

문자열에서 N개의 연속된 요소를 추출하는 통계 기반 언어 분석모델로, 문장 내 연속하여 동시 출현하는 단어와 그 빈도를 계산합니다. 문장을 몇 개의 단어 개수에 따라 나눌지에 따라 종류가 결정됩니다.

Bigram(2) Trigram(3) Tetragram(4)

상위 20개 단어 선택

16개 / 총 53816 개 ※ 전체 다운로드는 최대 20만 건까지 가능합니다. 선택 다운로드 전체 다운로드

<input type="checkbox"/>	단어1	단어2	빈도(건)
<input checked="" type="checkbox"/>	전세사기	피해자	1084
<input checked="" type="checkbox"/>	전세	보증금	740
<input checked="" type="checkbox"/>	전세사기	피해	677
<input checked="" type="checkbox"/>	전세사기	변호사	582
<input checked="" type="checkbox"/>	보증금	반환	490
<input checked="" type="checkbox"/>	전세	계약	477

선택 단어 시각화 >>

1 시각화 결과 크게보기 다운로드

2 시각화 설정

▶ 색상선택

선택상 ● 화살표 모양 ① ② ③ Tt 글자 크기 14

- ① **시각화 결과**
 N-gram 시각화 결과는 방향성이 있는 네트워크 그래프로 제공됩니다.
 Tip. 분석 목적에 따라 관련성 높은 단어를 선택하는 것이 중요하며, 정보를 제공하지 않는 일반적인 단어는 제외하는 것이 좋습니다.

- ② **시각화 설정**
 사용자는 시각화 결과의 속성을 원하는 대로 변경할 수 있습니다.

- ③ **크게보기**
 [크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
 Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.

- ④ **시각화 다운로드**
 시각화 결과를 PNG파일로 다운로드 합니다.

단어분석

TF-IDF 분석 결과를 확인합니다.

TEXTOM
백데이터팀님 | 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

단어 빈도
N-gram
TF-IDF
시계열

TF-IDF 1 분석 알고리즘

전체 문서에서 특정 단어가 언급된 횟수를 계산합니다.
단어 언급빈도는 중요도 분석을 통해 문서의 주제 또는 태도나 감성을 추론하는 지표로 활용될 수 있습니다.

상위 20개 | 단어 선택

8개 선택 / 총 6295 개 4 선택 다운로드 | 전체 다운로드

	단어	TF-IDF	DF	IDF
<input type="checkbox"/>	피해자	2170	881.000	1.308
<input type="checkbox"/>	변호사	2007	593.000	1.704
<input type="checkbox"/>	보증금	1823	1041.000	1.141
<input type="checkbox"/>	피해	1739	941.000	1.242
<input type="checkbox"/>	계약	1699	754.000	1.464
<input type="checkbox"/>	사기	1660	1105.000	1.082
<input type="checkbox"/>	지원	1625	410.000	2.073

5 선택단어 시각화 >>

시각화 결과 크게보기 | 다운로드

워드클라우드
버블차트
세로막대그래프
가로막대그래프

6

원쪽 리스트에서 시각화 할 단어를 선택해주세요.

시각화 설정
▶ 모양 선택

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **상위 단어 자동 선택**
전체 단어 중 상위 20개, 50개, 75개, 100개 단어를 자동으로 선택합니다.
- ③ **TF-IDF 단어 선택**
최상단의 체크박스를 클릭하면, 상위 100개의 단어를 자동으로 선택합니다. 단어를 직접 선택하려면 해당 단어 옆의 체크박스를 사용하세요.
Tip. TF-IDF 값이 높을수록 특정 문서 내에서 해당 단어가 얼마나 중요한지를 나타냅니다.
- ④ **분석 결과 다운로드**
[선택 다운로드], [전체 다운로드]를 클릭하면, 선택한 단어 또는 전체 분석 결과를 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.
- ⑤ 선택한 단어를 시각화 합니다.
- ⑥ 시각화 결과를 나타냅니다.

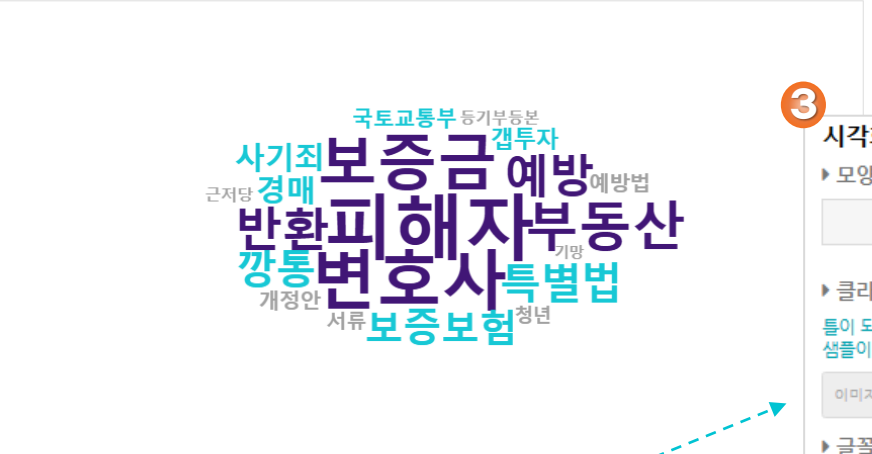
단어분석

TF-IDF 분석 결과를 시각화 합니다.

- <27> -

시각화 결과 크게보기 다운로드

1 **워드클라우드** 버블차트 세로막대그래프 가로막대그래프



2 **시각화 설정**

▶ 모양 선택

모양 선택 **직접 입력**

기본

▶ 글꼴 선택

기본

▶ 크기 선택

기본 기본 **적용**

▶ 색상선택

상 중 하

3 **시각화 설정**

▶ 모양 선택

모양 선택 **직접 입력**

▶ 클라우드 모양 변경 샘플 이미지1, 샘플 이미지2

틀이 되는 이미지를 업로드하여 특정 모양으로 시각화를 생성할 수 있습니다.
샘플 이미지를 참고하여 원하는 모양의 이미지(jpg, png)를 업로드해 보세요.

이미지 파일을 업로드해 보세요. **적용**

▶ 글꼴 선택

기본

▶ 크기 선택

기본 기본 **적용**

▶ 색상선택

상 중 하

4

5

- ① **시각화 결과**
TF-IDF 분석은 총 4종의 시각화 결과를 제공합니다.

워드클라우드	버블차트
가로막대차트	세로막대차트

- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
Tip. 워드클라우드 : 모양, 글꼴, 색상
버블차트 : 배경색, 글자색, 글자크기
세로·가로막대그래프 : 바 색상, 범례위치

- ③ **모양 직접 선택 (워드 클라우드)**
워드 클라우드 사용 시, 특정 모양으로 시각화를 생성할 수 있습니다.
Tip. 샘플 이미지를 참고하여 원하는 모양의 이미지를 업로드하세요.

- ④ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.

- ⑤ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

단어분석

수집량 시계열 분석을 진행합니다.

- <28> -

TEXTOM
백데이터팀님 + 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

단어 빈도
N-gram
TF-IDF
시계열

수집량
단어 빈도

수집량 시계열 분석

데이터 수집량을 기준으로 시계열 분석을 실시합니다.

② 기간단위 선택

일
주
월
년

③ 추세선

없음

① 분석 알고리즘

④ 적용

⑥ 다운로드

날짜	네이버(블로그)	네이버(뉴스)
2024-03-01	987	117
2024-04-01	954	119
2024-05-01	968	114

시각화 결과

크게보기
다운로드

시각화 설정

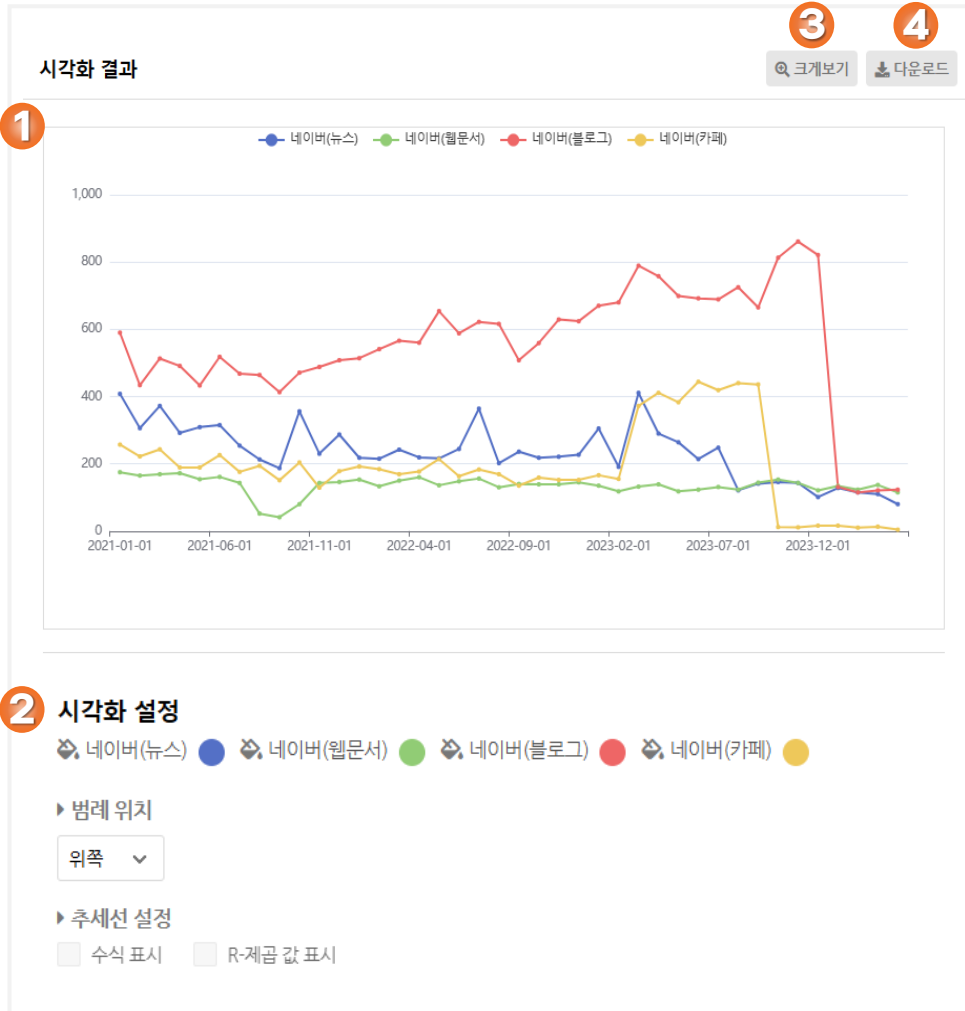
● 네이버(블로그)
 ● 네이버(뉴스)

▶ 범례 위치

- ① **분석 알고리즘**
버튼을 클릭하면 선형 추세선과 R-제곱 값의 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **기간단위 선택**
데이터 수집일을 기준으로 기간단위를 일, 주, 월, 년으로 설정할 수 있습니다.
Tip. 데이터 수집 단계에서 '단위수집' 기능을 사용하였을 때 선택 가능합니다.
'일' 선택 시 - 일, 주, 월, 년 사용가능
'주' 선택 시 - 주, 월, 년 사용가능
'월' 선택 시 - 월, 년 사용가능
'년' 선택 시 - 년 사용가능
- ③ **추세선 표시**
시각화 결과에 데이터 수집량의 추세선을 선형으로 표시하여 경향을 쉽게 파악할 수 있도록 합니다.
Tip. 시각화 결과에서 추세선을 표시하려면 반드시 [선형]으로 선택하세요.
- ④ 설정한 옵션을 시계열 분석에 적용합니다.
- ⑤ 수집량 시계열 분석 결과를 보여줍니다.
- ⑥ 분석 결과를 다운로드합니다.

단어분석

시계열 분석 - 데이터 수집량 분석 결과를 시각화 합니다.



- ① **시각화 결과**
분석 결과를 선형 그래프로 보여줍니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
Tip. 추세선을 설정하려면 분석 옵션에서 [추세선]을 [선형]으로 선택하세요.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG파일로 다운로드 합니다.

단어분석

단어빈도(TF) 시계열 분석을 진행합니다.

- <30> -

1 분석 알고리즘

데이터 수집량을 기준으로 시계열 분석을 실시합니다.

2 기간단위 선택

3 분석단어 선택

4 추세선

5 적용

6

날짜	특별법	보증보험	처벌
2024-03-01	80	164	81
2024-04-01	161	128	62
2024-05-01	241	145	53

7 다운로드

시각화 결과

분석단어선택 (시계열분석)

데이터명	생성날짜	용량
전세사기	2024-05-27	1.53 MB

단어빈도

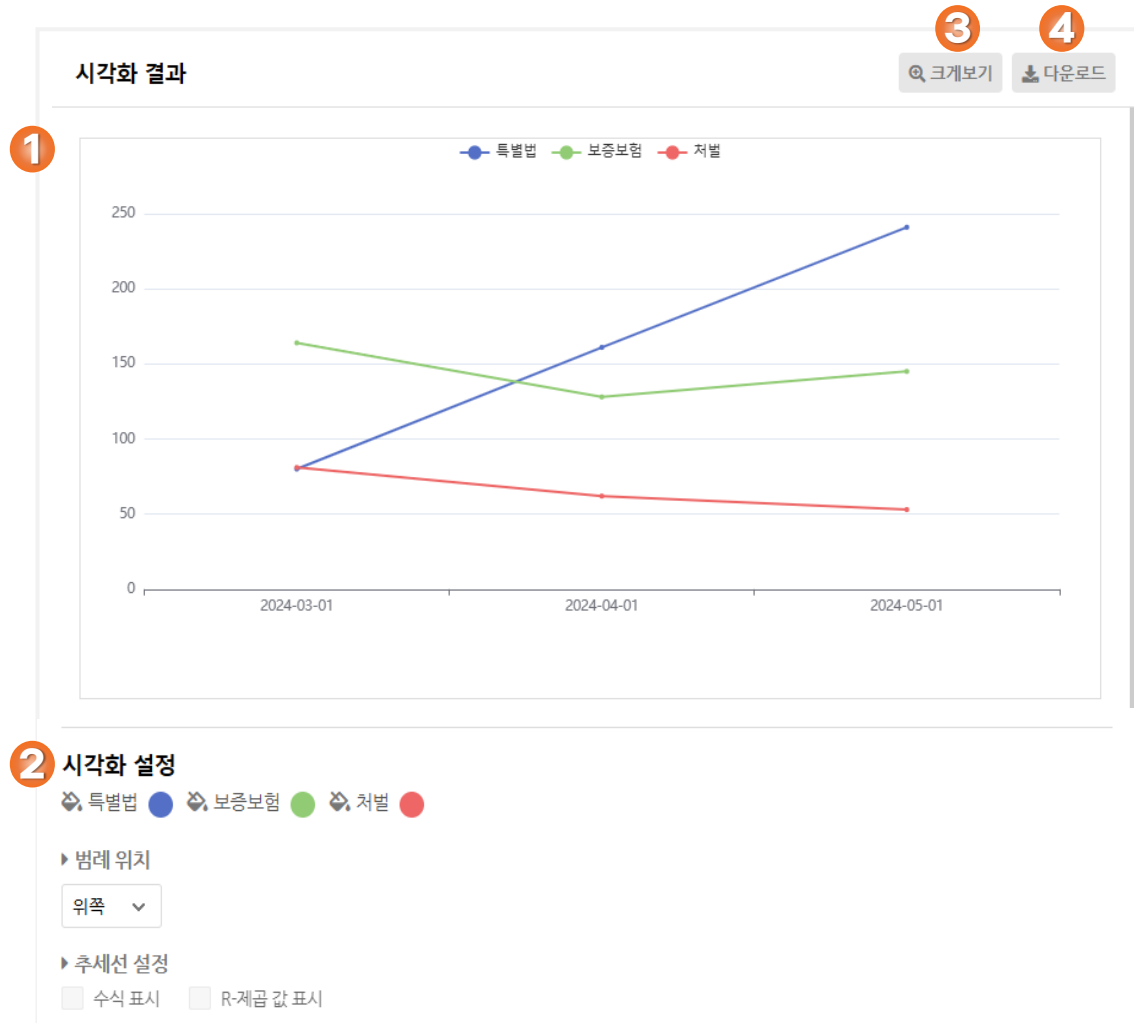
단어빈도 TF-IDF

선택단어수 확인 선택단어누적비율 %

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

단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/> 전세사기	7436	6.378945020631203	6.378945020631203
<input type="checkbox"/> 전세	4085	3.5043020991498746	9.883247119781078
<input type="checkbox"/> 피해자	1659	1.4231669969374887	11.306414116718567
<input type="checkbox"/> 보증금	1598	1.3708383731802936	12.67725248989886
<input type="checkbox"/> 사기	1535	1.3167940568408953	13.994046546739757
<input type="checkbox"/> 피해	1400	1.2009848075421845	15.19503135428194
<input type="checkbox"/> 변호사	1178	1.0105429309176381	16.205574285199578
<input type="checkbox"/> 계약	1161	0.9959595439689116	17.20153382916849
<input type="checkbox"/> 수	1123	0.9633613849070524	18.164895214075543

- 1 분석 알고리즘**
버튼을 클릭하면 선행 추세선과 R-제곱 값의 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- 2 기간단위 선택**
데이터 수집일을 기준으로 기간단위를 일, 주, 월, 년으로 설정할 수 있습니다.
Tip. 데이터 수집 단계에서 '단위수집' 기능을 사용하였을 때 선택 가능합니다.
'일' 선택 시 - 일, 주, 월, 년 사용가능
'주' 선택 시 - 주, 월, 년 사용가능
'월' 선택 시 - 월, 년 사용가능
'년' 선택 시 - 년 사용가능
- 3 분석단어 선택**
[바로 선택하기] 버튼을 클릭하여 2~10개 단어를 선택할 수 있습니다.
- 4 추세선 표시**
단어 빈도 시계열 분석에서 시각화 결과에 추세선을 표시하여 경향을 쉽게 파악할 수 있습니다. 추세선은 선행으로 표시됩니다.
Tip. 시각화 결과에서 추세선을 표시하려면 반드시 [선행]으로 선택하세요.
- 5 설정한 옵션을 시계열 분석에 적용합니다.**
- 6 단어 빈도 시계열 분석 결과를 보여줍니다.**
- 7 분석 결과를 엑셀(.xlsx) 파일로 다운로드합니다.**

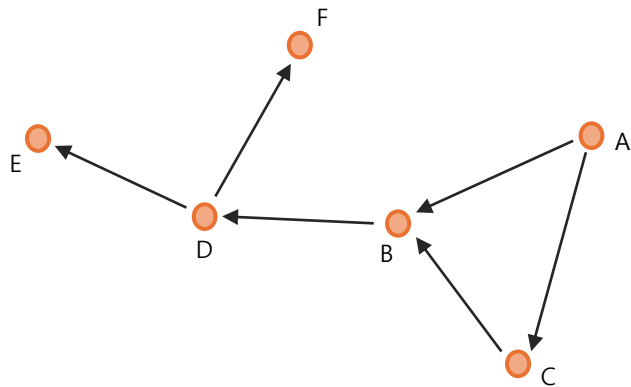


- ① **시각화 결과**
분석 결과를 선형 그래프로 보여줍니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
Tip. 추세선을 설정하려면 분석 옵션에서 [추세선]을 [선형]으로 선택하세요.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG파일로 다운로드 합니다.

매트릭스

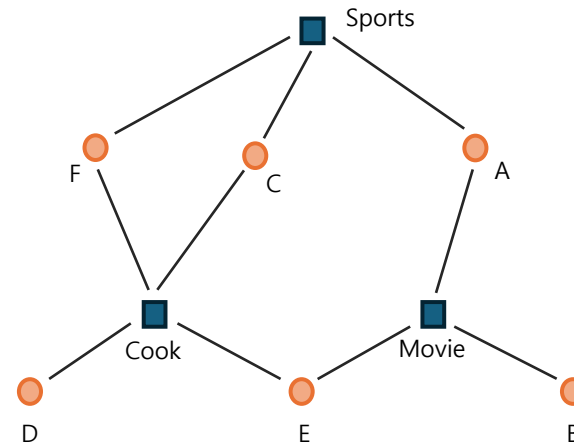
매트릭스 분석 방법에 대한 설명입니다.

1 1-Mode 분석 방법



	A	B	C	D	E	F
A	0	1	1	0	0	0
B	0	0	0	1	0	0
C	0	1	0	0	0	0
D	0	0	0	0	1	1
E	0	0	0	0	0	0
F	0	0	0	0	0	0

1 2-Mode 분석 방법



	Sports	Movie	Cook
A	1	1	0
B	0	1	0
C	1	0	1
D	0	0	1
E	0	1	1
F	1	0	1

매트릭스 데이터는 단어 간 공출현 빈도수로 작성된 매트릭스를 통해 단어들 간의 관계를 분석합니다.

- ① **1-Mode 매트릭스**
 모든 네트워크의 가장 일반적인 유형으로, 행과 열이 같은 단어로 이루어진 매트릭스입니다.
 Tip. 단어(메인노드) 간 관계를 분석할 때 적합합니다.
- ② **2-Mode 매트릭스**
 행과 열이 다른 단어로 이루어진 매트릭스로 서로 다른 유형의 조직 간 관계를 살펴볼 때 사용합니다.
 Tip. 단어(메인노드)와 범주(서브노드) 간의 관계를 분석할 때 적합합니다.

매트릭스

1-Mode 매트릭스 분석을 진행할 단어를 선택합니다.

TEXTOM
빅데이터팀님 | 데이터 추가

단어분석 | 매트릭스 | 네트워크 분석 | 토픽 분석 | 감성분석 | QAP 가설검정

1-Mode
2-Mode

매트릭스(Matrix Word) 1 분석 알고리즘

선택한 단어 간의 공출현(Co-occurrence) 매트릭스를 생성합니다.
※ 2개~200개 사이의 단어를 선택할 수 있습니다.

2 단어선택 [단어 선택하기] [파일 업로드]

3 상위 20개 단어 선택

4 0개 선택 / 총 6294 개

	단어	빈도(건)	백분율(%)	누적백분율(%)
<input type="checkbox"/>	전세사기	7436	6.379	6.379
<input type="checkbox"/>	전세	4085	3.504	9.883
<input type="checkbox"/>	피해자	1659	1.423	11.306
<input type="checkbox"/>	보증금	1598	1.371	12.677
<input type="checkbox"/>	사기	1535	1.317	13.994
<input type="checkbox"/>	피해	1400	1.201	15.195

6 ✓ 적용

매트릭스 결과 다운로드

매트릭스
옛지리스트
유클리디언 계수
코사인 계수
자카드 계수
상관 계수

7

왼쪽 리스트에서 분석할 단어를 선택해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **단어 선택 방법 설정**
분석에 사용할 2~200개의 단어를 선택합니다. 분석 콘솔에서 직접 선택하거나 엑셀 파일을 업로드할 수 있습니다.
- ③ **상위 단어 자동 선택**
전체 단어 중 상위 20개, 50개, 75개, 100개 단어를 자동으로 선택합니다.
- ④ **단어 선택하기**
최상단의 체크박스를 클릭하면, 상위 100개의 단어를 자동으로 선택합니다. 단어를 직접 선택하려면 해당 단어 옆의 체크박스를 사용하세요.
Tip. 분석 주제와 관련이 있는 핵심단어를 선택하는 것이 가장 중요합니다. 초기 선택 후 결과를 검토하고, 필요에 따라 조정하여 분석의 정확도를 높이세요.
- ⑤ **단어 목록 다운로드**
[선택 다운로드], [전체 다운로드]를 클릭하면, 선택한 단어 또는 전체 분석 결과를 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.
- ⑥ 선택한 단어를 분석에 적용합니다.
- ⑦ 분석 결과를 나타냅니다.

매트릭스

1-Mode 매트릭스 분석을 진행할 단어 목록을 파일로 업로드합니다.

1-Mode
2-Mode

매트릭스(Matrix Word) 1 분석 알고리즘

선택한 단어 간의 공출현(Co-occurrence) 매트릭스를 생성합니다.
※ 2개~200개 사이의 단어를 선택할 수 있습니다.

▶ 단어선택

단어 선택하기
파일 업로드

1 매트릭스 단어 파일을 업로드 해주세요. ※ xlsx 형식만 가능

2
예시파일

전체 단어 중, 분석 적용된 단어의 데이터를 확인할 수 있습니다.

4	번호	단어	빈도(건)
	1	전세	4085
	2	피해자	1659
	3	보증금	1598
	4	사기	1535
	5	피해	1400
	6	계약	1161

5
다운로드

3
✓ 적용

	A	B	C
1	전세		
2	피해자		
3	보증금		
4	사기		
5	피해		
6	계약		
7	주택		
8	임대인		
9	임차인		
10	보증보험		
11	세입자		
12	변호사		
13	소송		
14	특별법		

- ① **파일 업로드**
매트릭스 단어 목록이 작성된 엑셀 파일(.xlsx)을 업로드합니다.
- ② **예시파일**
해당 버튼을 클릭하면 예시 형식을 다운로드할 수 있습니다.
- ③ 선택한 단어를 분석에 적용합니다.
- ④ **선택 단어 확인**
전체 단어 중 분석 적용된 단어의 목록을 확인할 수 있습니다.
- ⑤ **단어 목록 다운로드**
분석 적용된 단어 목록을 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.

매트릭스

1-Mode 매트릭스 분석 결과를 확인합니다.

TEXTOM
데이터 분석 콘솔
빅데이터팀님
데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

1-Mode
2-Mode

매트릭스(Matrix Word) 분석 알고리즘

선택한 단어 간의 공출현(Co-occurrence) 매트릭스를 생성합니다.
※ 2개~200개 사이의 단어를 선택할 수 있습니다.

단어선택
단어 선택하기
파일 업로드

상위 20개 단어 선택

0개 선택 / 총 6294 개 선택 다운로드 전체 다운로드

	단어	빈도(건)	백분율(%)	누적백분율(%)
<input type="checkbox"/>	전세사기	7436	6.379	6.379
<input type="checkbox"/>	전세	4085	3.504	9.883
<input type="checkbox"/>	피해자	1659	1.423	11.306
<input type="checkbox"/>	보증금	1598	1.371	12.677
<input type="checkbox"/>	사기	1535	1.317	13.994
<input type="checkbox"/>	피해	1400	1.201	15.195

✓ 적용

매트릭스 결과 다운로드

매트릭스
옛지리스트
유클리디언 계수
코사인 계수
자카드 계수
상관 계수

	전세	피해자	보증금	사기	피해	계약	주택	임대인	임차인	보증보험	세입자
전세	0	1418	2867	1823	1613	2230	1217	896	722	1239	483
피해자	1418	0	525	583	1013	240	501	165	190	86	58
보증금	2867	525	0	778	557	631	374	506	390	339	257
사기	1823	583	778	0	751	640	250	276	187	168	161
피해	1613	1013	557	751	0	404	396	190	225	99	94
계약	2230	240	631	640	404	0	300	393	240	223	157
주택	1217	501	374	250	396	300	0	122	134	145	58
임대인	896	165	506	276	190	393	122	0	227	139	97
임차인	722	190	390	187	225	240	134	227	0	92	29
보증보험	1239	86	339	168	99	223	145	139	92	0	50
세입자	483	58	257	161	94	157	58	97	29	50	0

① **매트릭스 결과**
 분석 결과는 아래의 6가지 유사도 계수로 제공됩니다. 분석 명칭을 클릭하면, 해당 결과가 하단에 표시됩니다.

- 매트릭스
- 옛지리스트
- 유클리디언 계수
- 코사인 계수
- 자카드 계수
- 상관 계수

② **분석 결과 다운로드**
 선택한 분석 방법의 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
 Tip. 모든 분석 결과를 한 번에 다운로드 하는 기능은 제공하지 않습니다.

매트릭스

2-Mode 매트릭스 분석을 진행할 단어를 선택합니다.

- <37> -

1-Mode
2-Mode

매트릭스(Matrix Word) 1 분석 알고리즘

선택한 단어 간의 공출현(Co-occurrence) 매트릭스를 생성합니다.
※ 2개~200개 사이의 단어를 선택할 수 있습니다.

2 단어선택
☰ 단어 선택하기
📄 파일 업로드

※ 열단어, 행단어 모두 선택 완료 후 적용 버튼을 클릭하여 매트릭스 결과를 확인해 주세요.

열단어

4 상위 20개 단어 선택

행단어

5 0개 선택 / 총 6294 개 6 선택 다운로드 전체 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
<input type="checkbox"/> 전세사기	7436	6.379	6.379
<input type="checkbox"/> 전세	4085	3.504	9.883
<input type="checkbox"/> 피해자	1659	1.423	11.306
<input type="checkbox"/> 보증금	1598	1.371	12.677
<input type="checkbox"/> 사기	1535	1.317	13.994
<input type="checkbox"/> 피해	1400	1.201	15.195
<input type="checkbox"/> 변호사	1178	1.011	16.206
<input type="checkbox"/> 계약	1161	0.996	17.202
<input type="checkbox"/> 수	1123	0.963	18.165
<input type="checkbox"/> 것	1105	0.948	19.113
<input type="checkbox"/> 부동산	1001	0.859	19.972
<input type="checkbox"/> 등	889	0.763	20.734

7 열단어 선택 완료

8 적용

열단어
행단어

상위 20개 단어 선택

0개 선택 / 총 6294 개 선택 다운로드 전체 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
<input type="checkbox"/> 전세사기	7436	6.379	6.379
<input type="checkbox"/> 전세	4085	3.504	9.883
<input type="checkbox"/> 피해자	1659	1.423	11.306
<input type="checkbox"/> 보증금	1598	1.371	12.677
<input type="checkbox"/> 사기	1535	1.317	13.994
<input type="checkbox"/> 피해	1400	1.201	15.195
<input type="checkbox"/> 변호사	1178	1.011	16.206
<input type="checkbox"/> 계약	1161	0.996	17.202
<input type="checkbox"/> 수	1123	0.963	18.165
<input type="checkbox"/> 것	1105	0.948	19.113
<input type="checkbox"/> 부동산	1001	0.859	19.972
<input type="checkbox"/> 등	889	0.763	20.734

행단어 선택 완료

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **단어 선택 방법 설정**
분석에 사용할 2~200개의 단어를 선택합니다. 분석 콘솔에서 직접 선택하거나 엑셀 파일을 업로드할 수 있습니다.
- ③ 탭을 선택하여 열단어와 행단어로 이동할 수 있습니다.
- ④ **상위 단어 자동 선택**
전체 단어 중 상위 20개, 50개, 75개, 100개 단어를 자동으로 선택합니다.
- ⑤ **단어 선택하기**
최상단의 체크박스를 클릭하면, 상위 100개의 단어를 자동으로 선택합니다. 단어를 직접 선택하려면 해당 단어 옆의 체크박스를 사용하세요.
Tip. 분석 주제와 관련이 있는 핵심단어를 선택하는 것이 가장 중요합니다. 초기 선택 후 결과를 검토하고, 필요에 따라 조정하여 분석의 정확도를 높이세요.
- ⑥ **단어 목록 다운로드**
[선택 다운로드], [전체 다운로드]를 클릭하면, 선택한 단어 또는 전체 분석 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
- ⑦ 열단어와 행단어 선택 후 완료 버튼을 눌러 저장합니다.
- ⑧ 선택한 단어를 분석에 적용합니다.

매트릭스

2-Mode 매트릭스 분석 결과를 확인합니다.

- <38> -

1-Mode
2-Mode

매트릭스(Matrix Word)

선택한 단어 간의 공출현(Co-occurrence) 매트릭스를 생성합니다.
※ 2개~200개 사이의 단어를 선택할 수 있습니다.

단어선택 📄 단어 선택하기 📄 파일 업로드

열단어 C:\fakepath\2-mode_열.xlsx 👤 예시파일

행단어 C:\fakepath\2-mode_행.xlsx

전체 단어 중, 분석 적용된 단어의 데이터를 확인할 수 있습니다. 📄 다운로드

번호	단어	빈도(건)
1	임대인	595
2	임차인	454
3	보증보험	438
4	부동산	1001

✓ 적용

매트릭스 결과

매트릭스
옛지리스트
유클리디언 계수
코사인 계수
자카드 계수
상관 계수

	임대인	임차인	보증보험	부동산	변호사	국토교통부	신혼부부	초년생
보증금	506	390	339	437	703	36	33	23
사기	276	187	168	592	757	14	16	29
피해자	165	190	86	194	322	234	23	28
특별법	24	29	18	36	13	56	2	2
주택	122	134	145	159	110	61	36	14
지원	25	158	60	61	52	171	22	8
반환	232	179	294	191	424	24	19	15
예방	150	107	264	236	40	30	9	11
계약	393	240	223	498	314	31	18	26
강통	114	89	37	168	126	6	6	21
경매	42	61	8	68	34	9	2	1
갹투자	53	30	11	72	34	1	2	3

① 매트릭스 결과
분석 결과는 아래의 6가지 유사도 계수로 제공됩니다. 분석 명칭을 클릭하면, 해당 결과가 하단에 표시됩니다.

- 매트릭스
- 옛지리스트
- 유클리디언 계수
- 코사인 계수
- 자카드 계수
- 상관 계수

② 분석 결과 다운로드
선택한 분석 방법의 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
Tip. 모든 분석 결과를 한 번에 다운로드 하는 기능은 제공하지 않습니다.

매트릭스

2-Mode 매트릭스 분석 결과입니다.

1 매트릭스

매트릭스	워드리스트	유클리디언 계수	코사인 계수	자카드 계수	상관 계수			
보통글	506	390	339	437	703	36	33	23
사기	276	187	168	592	757	14	16	29
피해자	165	190	86	194	322	234	23	28
특별법	24	29	18	36	13	56	2	2
주석	122	134	145	159	110	61	36	14
지침	25	158	60	61	52	171	22	8
반환	232	179	294	191	424	24	19	15
예방	150	107	264	236	40	30	9	11
계약	393	240	223	498	314	31	18	26
강릉	114	89	37	168	126	6	6	21
경매	42	61	8	68	34	9	2	1
건투자	53	30	11	72	34	1	2	3

2 매트릭스

워드1	워드2	Weight(Tf)
일반인	보통글	506
일반인	사기	276
일반인	피해자	165
일반인	특별법	24
일반인	주석	122
일반인	지침	25
일반인	반환	232
일반인	예방	150
일반인	계약	393
일반인	강릉	114
일반인	경매	42
일반인	건투자	53

3 매트릭스

매트릭스	워드리스트	유클리디언 계수	코사인 계수	자카드 계수	상관 계수			
보통글	0.9448	0.9311	0.9418	0.9252	0.946	0.7226	0.7705	0.7226
사기	0.9009	0.8639	0.8915	0.948	0.9588	0.2928	0.5917	0.6666
피해자	0.8951	0.9191	0.8829	0.8994	0.8852	0.9385	0.6837	0.6464
특별법	0.6666	0.6464	0.6837	0.7817	0.5327	0.7958	0	0
주석	0.874	0.8509	0.9037	0.9046	0.8874	0.8038	0.7817	0.8356
지침	0.7642	0.9239	0.9019	0.8377	0.9019	0.9169	0.7958	0.4226
반환	0.9024	0.8957	0.9166	0.9019	0.9357	0.6464	0.6984	0.7226
예방	0.8804	0.8686	0.9105	0.9	0.7958	0.6464	0.6464	0.4226
계약	0.9426	0.9178	0.9291	0.9489	0.9324	0.7763	0.622	0.7418
강릉	0.8837	0.8585	0.7763	0.9067	0.8867	0	0	0.7914
경매	0.5917	0.8333	0	0.8174	0.7226	0	0	0
건투자	0.7958	0.7226	0.4226	0.8309	0.8398	0	0	0

4 매트릭스

매트릭스	워드리스트	유클리디언 계수	코사인 계수	자카드 계수	상관 계수			
보통글	0.079	0.0835	0.0427	0.0619	0.0595	0.0159	0.0271	0.0215
사기	0.0564	0.0545	0.0286	0.0831	0.0694	0.0099	0.0178	0.0318
피해자	0.0256	0.0342	0.0132	0.026	0.0206	0.0593	0.0158	0.0268
특별법	0.0103	0.0173	0.0073	0.0098	0.0045	0.0507	0.0026	0.0026
주석	0.0311	0.0406	0.022	0.0335	0.0146	0.0445	0.0156	0.0139
지침	0.0055	0.031	0.0104	0.0121	0.0058	0.0802	0.0185	0.0094
반환	0.0612	0.0602	0.0521	0.0443	0.055	0.0191	0.0275	0.0187
예방	0.0426	0.0441	0.0489	0.0537	0.0105	0.0256	0.0121	0.0198
계약	0.07	0.0645	0.0391	0.0695	0.0437	0.0146	0.0229	0.0341
강릉	0.0405	0.0397	0.0161	0.0461	0.0316	0.0076	0.0137	0.0252
경매	0.021	0.0291	0.0059	0.0294	0.0081	0.018	0.0072	0.0036
건투자	0.0295	0.0238	0.0091	0.0281	0.016	0.0027	0.0097	0.0148

5 매트릭스

매트릭스	워드리스트	유클리디언 계수	코사인 계수	자카드 계수	상관 계수			
보통글	0	0	0	0	0	0	0	0
사기	0	0	0	0	0	0	0	0
피해자	0	0	0	0	0	0	0	0
특별법	0	0	0	0	0	0	0	0
주석	0	0	0	0	0	0	0	0
지침	0	0	0	0	0	0	0	0
반환	0	0	0	0	0	0	0	0
예방	0	0	0	0	0	0	0	0
계약	0	0	0	0	0	0	0	0
강릉	0	0	0	0	0	0	0	0
경매	0	0	0	0	0	0	0	0
건투자	0	0	0	0	0	0	0	0

6 매트릭스

매트릭스	워드리스트	유클리디언 계수	코사인 계수	자카드 계수	상관 계수			
보통글	0.1404	0.1396	0.0846	0.0232	0.0472	0.0628	0.0189	0.014
사기	0.0004	0.0199	0.0248	0.0505	0.0733	0.0881	0.0181	0.0048
피해자	0.0762	0.0285	0.0784	0.1252	0.088	0.1875	0.0158	0.0033
특별법	0.0651	0.0491	0.0486	0.0837	0.0938	0.0748	0.0211	0.0266
주석	0.0191	0.017	0.0293	0.0513	0.0799	0.0395	0.0661	0.0002
지침	0.0853	0.047	0.0382	0.0978	0.0966	0.2164	0.0281	0.0172
반환	0.0873	0.0874	0.1837	0.027	0.0798	0.0277	0.0262	0.0055
예방	0.0318	0.0219	0.1776	0.0344	0.1085	0.0039	0.0035	0.002
계약	0.1326	0.0737	0.0468	0.0718	0.0452	0.0429	0.0006	0.0157
강릉	0.0165	0.025	0.0325	0.013	0.0331	0.0455	0.0099	0.0409
경매	0.0145	0.0275	0.0467	0.0218	0.0602	0.0169	0.0134	0.0236
건투자	0.0334	0.0141	0.0214	0.0205	0.0274	0.0298	0.0103	0.0055

- 매트릭스**
전체 문서 내에서 선택한 단어 간 공출현 횟수를 나타냅니다.
- 워드리스트**
선택 단어 쌍(word1, word2)과 그 빈도(Weight)를 리스트 형태로 보여줍니다.
- 유클리디언 계수**
선택한 단어 간의 유사성을 유클리디언 거리로 측정하여 제공합니다.
Tip. 두 벡터간 값이 작을수록 유사합니다.
- 코사인 계수**
두 단어 벡터 간의 코사인 유사도를 계산하여 제공합니다.
Tip. 계수의 범위는 0에서 1사이로, 1에 가까울수록 두 벡터의 유사성이 높음을 의미합니다.
- 자카드 계수**
단어 간 자카드 유사도를 보여줍니다.
Tip. 계수의 범위는 0에서 1까지로, 값이 클수록 유사성이 높음을 의미합니다.
- 상관 계수**
단어 간 상관 관계를 나타냅니다.
Tip. 값이 클수록 두 변수 간의 관계가 강함을 나타냅니다.

TEXTOM 홈 데이터 분석 콘솔 백데이터팀 데이터 추가

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 중심성 에고네트워크 최단경로 CONCOR 클러스터링

네트워크 속성 1 분석 알고리즘

네트워크 전체에 대한 구조적 기술 통계량을 분석합니다. 네트워크의 노드들의 상호연결성으로 관련 지수로 측정합니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 분석단어 2 바로 선택하기

네트워크 속성 분석 결과

3 ☰

매트릭스 분석 단어를 선택해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **분석단어 선택**
[바로 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동합니다.
Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.
- ③ **분석 결과를 나타냅니다.**

네트워크 분석

네트워크 속성 분석을 진행할 단어를 선택하고 결과를 확인합니다.

TEXTOM
데이터 분석 콘솔
백데이터입님
데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성
중심성
애고네트워크
최단경로
CONCOR
클러스터링

네트워크 속성

네트워크 전체에 대한 구조적 기술 통계량을 분석합니다. 네트워크의 노드들의 상호연결성으로 관련 지수로 측정합니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 분석단어

단어	빈도(건)	백분율(%)	누적백분율(%)
주택	806	0.691	21.426
반환	758	0.65	23.425
예방	649	0.557	25.192
임대인	595	0.51	25.703
강통	505	0.433	29.467
소송	484	0.415	29.882
특별법	482	0.413	30.296
임차인	454	0.389	31.887
부조부하	438	0.376	32.037

1
« 다시 선택하기

네트워크 속성 분석 결과

3

Network Measures	Value
Nodes	67
Total Edges	1079
Diameter	2
Degree Centralization	0.4965
Closeness Centralization	0.60001
Betweenness Centralization	1.0E-5
Connected Components	1
Overall Reciprocity	0

4
다운로드

- ① **단어 다시 선택하기**
[다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.
- ② **분석 단어 다운로드**
선택한 단어를 엑셀(.xlsx) 파일로 다운로드합니다.

- ③ **네트워크 속성 분석 결과**
1-Mode 매트릭스에 대한 네트워크 구조적 속성 지표 8종을 산출합니다.

Nodes	노드의 총 개수
Total Edges	엣지의 총 개수
Diameter	직경
Degree Centralization	연결정도 집중도
Closeness Centralization	근접 집중도
Betweenness Centralization	매개집중도
Connected Components	연결 컴포넌트 수
Overall Reciprocity	상호성

- ④ **분석 결과 다운로드**
분석 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

TEXTOM 데이터 분석 콘솔 백데이터팀님 데이터 추가

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 **중심성** 애고네트워크 최단경로 CONCOR 클러스터링

중심성 1 분석 알고리즘

네트워크에서 특정 노드의 상대적 중요성을 수치화한 값입니다. 중심성 분석을 통해 구조적 위치의 관점에서 보다 영향력 있는(의미있는) 노드를 파악할 수 있습니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 분석단어 2 바로 선택하기

중심성 분석 결과

3 리스트 아이콘

매트릭스 분석 단어를 선택해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **분석단어 선택**
[바로 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동합니다.
Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.
- ③ **분석 결과를 나타냅니다.**

네트워크 분석

중심성 분석을 진행할 단어를 선택하고 결과를 확인합니다.

TEXTOM
데이터 분석 콘솔
백데이터입님
데이터추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성
중심성
예고네트워크
최단경로
CONCOR
클러스터링

중심성 분석 알고리즘

네트워크에서 특정 노드의 상대적 중요성을 수치화한 값입니다. 중심성 분석을 통해 구조적 위치의 관점에서 보다 영향력 있는(의미있는) 노드를 파악할 수 있습니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

분석단어 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
주택	806	0.691	21.426
반환	758	0.65	23.425
예방	649	0.557	25.192
임대인	595	0.51	25.703
강통	505	0.433	29.467
소송	484	0.415	29.882
특별법	482	0.413	30.296
임차인	454	0.389	31.887

1 << 다시 선택하기

중심성 분석 결과 다운로드

단어	연결 중심성	근접 중심성	매개 중심성	아이겐벡터 중심성	페이지 랭크	클러스터링 계수
주택	0.97	0.971	0.046	0.193	0.073	0.489
반환	0.924	0.93	0.041	0.188	0.07	0.508
예방	0.848	0.868	0.024	0.181	0.054	0.569
임대인	0.879	0.892	0.03	0.184	0.053	0.546
강통	0.864	0.88	0.028	0.182	0.043	0.549
소송	0.833	0.857	0.023	0.179	0.039	0.576
특별법	0.803	0.835	0.03	0.166	0.034	0.536
임차인	0.909	0.917	0.031	0.189	0.041	0.537
보증보험	0.803	0.835	0.022	0.172	0.041	0.575
빌라	0.818	0.846	0.024	0.175	0.035	0.57
아파트	0.742	0.795	0.018	0.164	0.025	0.618
경매	0.803	0.835	0.021	0.173	0.026	0.581
공인중개사	0.727	0.786	0.014	0.164	0.021	0.644
사기죄	0.652	0.742	0.009	0.152	0.021	0.682

① **단어 다시 선택하기**
[다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.

② **분석 단어 목록 다운로드**
분석 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

③ **중심성 분석 결과**
분석 결과로 중심성 지표 5종과 클러스터링 계수를 보여줍니다. 중심성 지표는 다음과 같습니다.

연결 중심성 (Degree Centrality)
근접 중심성 (Closeness Centrality)
매개 중심성 (Betweenness Centrality)
아이겐벡터 중심성 (Eigenvector Centrality)
페이지 랭크 (PageRank)

④ **분석 결과 다운로드**
분석 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

네트워크 분석

에고 네트워크 분석을 진행합니다.

- <44> -

TEXTOM 데이터 분석 콘솔 백데이터팀 데이터수거

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 중심성 **에고네트워크** 최단경로 CONCOR 클러스터링

1 분석 알고리즘

전체에서 에고(Ego) 노드를 대상으로 하는 네트워크를 추출하며, 에고 노드와 연결된 다른 노드(타자, Alter)와의 연결 관계를 분석합니다.
※ 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

2 에고(Ego) 노드

3 분석단어 바로 선택하기

에고네트워크 분석 결과

에고네트워크 속성 에고네트워크 중심성 시각화 결과

4 분석 옵션을 선택해주세요.

- 분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- 에고 노드 단어 입력**
분석의 중심이 될 단어를 입력합니다. 해당 단어를 중심으로 네트워크를 추출합니다.
- 분석 단어 선택**
에고 노드(단어)와 관련된 노드(단어)를 분석하기 진행할 단어를 선택합니다.
Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.
- 분석 결과를 나타냅니다.

네트워크 분석

예고 네트워크 분석을 진행할 단어를 선택하고 결과를 확인합니다.

TEXTOM
데이터 분석 콘솔
백데이터담당
데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성
중심성
예고네트워크
최단경로
CONCOR
클러스터링

예고네트워크 분석 알고리즘

전체에서 예고(Ego) 노드를 대상으로 하는 네트워크를 추출하여, 예고 노드와 연결된 다른 노드(타자, Alter)와의 연결 관계를 분석합니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 예고(Ego) 노드

1 분석단어

단어	빈도(건)	백분율(%)	누적백분율(%)
주택	806	0.691	21.426
반환	758	0.65	23.425
예방	649	0.557	25.192
임대인	595	0.51	25.703
강통	505	0.433	29.467
소송	484	0.415	29.882
특별법	482	0.413	30.296

2 다운로드

예고네트워크 분석 결과

예고네트워크 속성
예고네트워크 중심성
시각화 결과

예고(Ego)	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
특별법	53	792	2756	0.28737	1964	0.00002

다운로드

3 << 다시 선택하기
4 >> 적용

① 분석 단어 목록을 확인할 수 있습니다.

② 분석 단어 목록 다운로드
 분석결과로 추출된 토픽과 단어의 토픽확률(%)을 산출합니다.

③ 단어 다시 선택하기
 [다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.
 Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.

④ 예고 노드와 함께 분석할 단어를 분석에 적용합니다.

에고네트워크 분석 결과

1

에고네트워크 속성

에고네트워크 중심성

시각화 결과

다운로드

에고(Ego)	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
특별법	53	792	2756	0.28737	1964	0.00002

에고네트워크 속성

2

에고네트워크 중심성

시각화 결과

3 다운로드

에고(Ego)	타자(Alter)	연결정도	근접	매개	아이겐벡터	페이지 랭크	클러스터링 계수
	반환	0.925	0.93	0.037	0.192	0.079	0.571
	예방	0.868	0.883	0.022	0.189	0.047	0.633
	임대인	0.868	0.883	0.023	0.188	0.044	0.629
	강통	0.887	0.898	0.024	0.191	0.032	0.617
	소송	0.83	0.855	0.018	0.184	0.036	0.656

① **에고 네트워크 속성**
 분석 결과로 에고 네트워크 지표 6종을 제공합니다.

Size	에고 네트워크 크기
Ties	직접 연결된 엣지 수
Pairs	전체 노드 쌍의 수
Ego-Density	에고 네트워크 밀도
Broker	중개자
Ego Betweenness	에고 매개성

② **에고 네트워크 중심성**
 분석 결과로 중심성 지표 5종과 클러스터링 계수를 보여줍니다. 중심성 지표는 다음과 같습니다.

연결 중심성 (Degree Centrality)
근접 중심성 (Closeness Centrality)
매개 중심성 (Betweenness Centrality)
아이겐벡터 중심성 (Eigenvector Centrality)
페이지 랭크 (PageRank)

③ **분석 결과 다운로드**
 분석 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

네트워크 분석

예고 네트워크 분석 결과를 시각화 합니다.

- <47> -

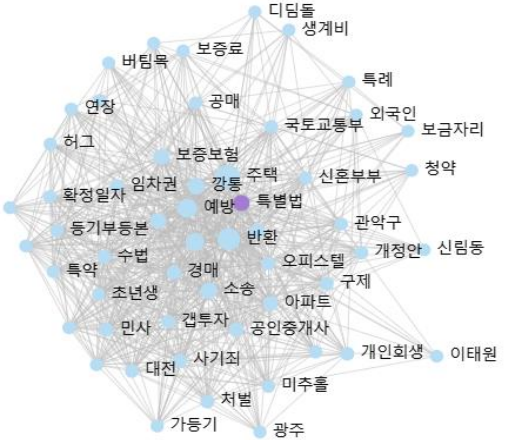
예고네트워크 속성
예고네트워크 중심성

시각화

3
4

🔍 크게보기
📄 다운로드

1



2

시각화 설정

▶ 색상선택

● 예고노드
 ● 타자노드
 ● 연결선

▶ 글자 설정

● 색상
 T 글자 크기

▶ 노드 설정

크기

- ① **시각화 결과**
 예고 네트워크의 시각화 결과는 방향성이 없는 네트워크 그래프로 제공됩니다.
 Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ② **시각화 설정**
 그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
 Tip. 단어 빈도를 기준으로 설정하면 자주 등장하는 단어가 더 큰 노드로 표시, 매개 중심성을 기준으로 하면 네트워크 내에서 중요한 역할을 하는 단어가 더 큰 노드로 표시됩니다.
- ③ **크게보기**
 [크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
 Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ④ **시각화 다운로드**
 시각화 결과를 PNG 파일로 다운로드 합니다.

TEXTOM ☰ 데이터 분석 콘솔 👤 백데이터팀님 ➕ 데이터 추가

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 중심성 에고네트워크 **최단경로** CONCOR 클러스터링

최단경로 1 📘 분석 알고리즘

경로(Path)란 한 노드에서 다른 노드에 도착하기 위해 거치는 연결 노드들을 의미합니다. 최단경로(Shortest path)는 네트워크 내에서 가장 길이가 짧은 경로를 의미합니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

2 ▶ 시작단어

▶ 도착단어

3 ▶ 분석단어

최단경로 분석 결과

4 ☰

분석 옵션을 선택해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② 네트워크 내에서 경로를 확인하고자 하는 최초 시작단어와 최종 도착단어를 입력합니다.
Tip. 반드시 네트워크 내에 존재하는 단어를 입력하세요.
- ③ **분석 단어 선택**
[바로 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동합니다.
Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.
- ④ 시각화 결과를 보여줍니다.

네트워크 분석

최단 경로 분석을 진행할 단어를 선택하고 결과를 확인합니다.

TEXTOM
백데이터팀님 > 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성

중심성

에고네트워크

최단경로

CONCOR

클러스터링

최단경로 ● 분석 알고리즘

경로(Path)란 한 노드에서 다른 노드에 도착하기 위해 거치는 연결 노드들을 의미합니다. 최단경로(Shortest path)는 네트워크 내에서 가장 길이가 짧은 경로를 의미합니다.
※ 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 시작단어:

▶ 도착단어:

1 분석단어

단어	빈도(건)	백분율(%)	누적백분율(%)
주택	806	0.691	21.426
반환	758	0.65	23.425
예방	649	0.557	25.192
임대인	595	0.51	25.703
강릉	505	0.433	29.467
수소	484	0.415	29.882

4 다운로드

최단경로 분석 결과

최단경로

시각화 결과

시작단어	도착단어	최단경로 길이	경로
확정일자	개정안	2	<div style="font-size: 0.8em; text-align: left; padding: 5px;"> <p>6 다운로드</p> 확정일자, 주택, 개정안 확정일자, 반환, 개정안 확정일자, 예방, 개정안 확정일자, 임대인, 개정안 확정일자, 강릉, 개정안 확정일자, 수소, 개정안 확정일자, 특별법, 개정안 확정일자, 임차인, 개정안 확정일자, 보증보험, 개정안 확정일자, 아파트, 개정안 확정일자, 경매, 개정안 확정일자, 공인중개사, 개정안 확정일자, 사기죄, 개정안 확정일자, 구제, 개정안 확정일자, 특약, 개정안 확정일자, 허그, 개정안 </div>

2 << 다시 선택하기

3 >> 적용

- ① 분석 단어 목록을 확인할 수 있습니다.
- ② **단어 다시 선택하기**
[다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.
- ③ 선택한 단어를 분석에 적용합니다.
- ④ **분석 단어 목록 다운로드**
분석 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
- ⑤ **최단거리 분석 결과**
분석 결과로 최단경로 길이와 세부 경로를 보여줍니다.
- ⑥ **분석 결과 다운로드**
군집 및 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

네트워크 분석

최단경로 분석 결과를 시각화 합니다.

최단경로
시각화 결과 3 4

🔍 크게보기
📄 다운로드

1

2

시각화 설정

▶ 색상선택

예고노드 ●
 타자노드 ●
 연결선 ●

▶ 글자 설정

색상 ●
 글자 크기

▶ 노드 설정

크기

- ① **시각화 결과**
 예고 네트워크의 시각화 결과는 방향성이 없는 네트워크 차트로 제공됩니다.
 Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ② **시각화 설정**
 그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
 Tip. 노드 크기 설정 시, 단어 빈도를 기준으로 설정하면 자주 등장하는 단어를, 매개 중심성을 기준으로 하면 네트워크 내에서 중요한 역할을 하는 단어를 더 크게 표시합니다.
- ③ **크게보기**
 [크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
 Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ④ **시각화 다운로드**
 시각화 결과를 PNG 파일로 다운로드 합니다.

TEXTOM 데이터 분석 콘솔 빅데이터팀님 데이터 추가

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 중심성 에고네트워크 최단경로 **CONCOR** 클러스터링

CONCOR 1 분석 알고리즘

동시에 출현하는 단어 사이의 관계를 분석하여 구조적 동위에 위치한 단어들을 군집화 하는 분석 기법입니다.
* 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

2 ▶ 군집 수 4

3 ▶ 분석단어 바로 선택하기

CONCOR 분석 결과

CONCOR 시각화 결과

4 분석 옵션을 선택해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **군집 수 선택**
군집 수를 선택합니다.
- ③ **분석 단어 선택**
[바로 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동합니다.
Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.
- ④ **분석 결과를 보여줍니다.**

네트워크 분석

CONCOR 분석 결과를 확인합니다.

TEXTOM
빅데이터팀님 > 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성

중심성

애고네트워크

최단경로

CONCOR

클러스터링

CONCOR 1 분석 알고리즘

동시에 출현하는 단어 사이의 관계를 분석하여 구조적 등위에 위치한 단어들을 군집화 하는 분석 기법입니다.
※ 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 군집 수

▶ 분석단어 4 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
오피스텔	194	0.166	47.441
갠투자	168	0.144	50.584
임차권	164	0.141	50.869
국토교통부	161	0.138	51.007
대전	153	0.131	51.943
예방법	144	0.124	53.075
개정안	141	0.121	53.809

2 << 다시 선택하기
3 >> 적용

CONCOR 분석 결과

CONCOR
시각화 결과

5	군집	단어
	1	구매
	1	연장
	1	마음
	1	초년생
	1	특약
	1	선순위
	1	내용증명
	1	대항력
	1	신혼부부
	1	미추홀
	1	강서구
	1	전월세

6 다운로드

- ① 분석 단어 목록을 확인할 수 있습니다.
- ② **단어 다시 선택하기**
[다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.
- ③ 선택한 단어를 분석에 적용합니다.
- ④ **분석 단어 목록 다운로드**
분석 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
- ⑤ **CONCOR 분석 결과**
분석을 진행할 단어를 선택하고 결과를 확인합니다.
- ⑥ **분석 결과 다운로드**
군집 및 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

CONCOR 분석 결과

CONCOR 시각화 결과 ③ ④

① 

② **시각화 설정**

▶ 색상선택

군집1
 군집2
 군집3
 군집4
 연결선

▶ 글자 설정

색상
 Tt 글자크기

▶ 노드 설정

크기

- ① **시각화 결과**
CONCOR 분석 결과는 네트워크 차트로 제공됩니다.
Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
Tip. 단어 빈도를 기준으로 설정하면 자주 등장하는 단어가 더 큰 노드로 표시, 매개 중심성을 기준으로 하면 네트워크 내에서 중요한 역할을 하는 단어가 더 큰 노드로 표시됩니다.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

네트워크 분석

클러스터링 분석을 진행합니다.

- <54> -

TEXTOM 데이터 분석 콘솔

단어분석 매트릭스 **네트워크 분석** 토픽 분석 감성분석 QAP 가설검정

네트워크 속성 중심성 에고네트워크 최단경로 CONCOR **클러스터링**

클러스터링 1 분석 알고리즘

네트워크 내의 소규모 그룹을 파악하기 위한 계층적 군집 분석(Hierarchical clustering) 방식입니다. 각 노드가 다른 노드들과 맺는 관계의 패턴을 계산하여 유사한 관계패턴을 갖는 노드들이 동일한 군집으로 분류됩니다.
 ※ 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

2 분석모델 루베인(Louvain) v

3 분석단어 바로 선택하기

클러스터링 분석 결과

클러스터링 시각화 결과

4 분석 옵션을 선택해주세요.

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

② **분석 모델 선택**
클러스터링 분석은 다음의 4종의 알고리즘을 제공합니다.

- 루베인 (Louvain)
- 레이든 (Leiden)
- GN (Girvan Newman)
- Clauset-Newman-Moore

Tip. 빠른 분석이 필요한 경우 : 루베인
 높은 정확도 : 레이든
 단순한 구조의 네트워크 분석 : GN
 계층적 구조 : Clauset-Newman-Moore

③ **분석 단어**
[바로 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동합니다.
 Tip. 반드시 1-Mode 매트릭스 분석이 선행되어야 합니다.

④ 분석 결과를 보여줍니다.

네트워크 분석

클러스터링 분석을 진행할 단어를 선택하고 결과를 확인합니다.

TEXTOM
백데이터팀님 | 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

네트워크 속성
중심성
예고네트워크
최단경로
CONCOR
클러스터링

클러스터링

네트워크 내의 소규모 그룹을 파악하기 위한 계층적 군집 분석(Hierarchical clustering) 방식입니다. 각 노드가 다른 노드들과 맺는 관계의 패턴을 계산하여 유사한 관계패턴을 갖는 노드들이 동일한 군집으로 분류됩니다.
※ 네트워크 분석은 [매트릭스] 1-Mode 선택 단어를 중심으로 산출됩니다.

▶ 분석모델: 루베인(Louvain) ▼

1 ▶ 분석단어

단어	빈도(건)	백분율(%)	누적백분율(%)
주택	806	0.691	21.426
반환	758	0.65	23.425
예방	649	0.557	25.192
임대인	595	0.51	25.703
강동	505	0.433	29.467
소송	484	0.415	29.882
특별법	482	0.413	30.296

4 다운로드

2 << 다시 선택하기 3 >> 적용

클러스터링 분석 결과

클러스터링
시각화 결과

군집	단어
1	주택
1	임대인
1	특별법
1	국토교통부
1	대전
1	개정안
1	구제
1	위원회
1	등기부등본
1	근저당
1	확정일자
1	대항력

6 다운로드

- ① 분석 단어 목록을 확인할 수 있습니다.
- ② **단어 다시 선택하기**
[다시 선택하기] 클릭 시, 1-Mode 매트릭스 단어 선택 페이지로 이동하여 분석할 단어를 재선택할 수 있습니다.
- ③ 선택한 단어를 분석에 적용합니다.
- ④ **분석 단어 목록 다운로드**
분석 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.
- ⑤ **클러스터링 분석 결과**
분석 결과로 단어와 그 단어가 속한 그룹을 산출하여 보여줍니다.
- ⑥ **분석 결과 다운로드**
군집 및 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

클러스터링 분석 결과

클러스터링 시각화 결과 **3** **4**

🔍 크게보기 📄 다운로드

1

2 **시각화 설정**

▶ 색상선택

군집1
 군집2
 군집3
 군집4
 군집5
 군집6
 연결선

▶ 글자 설정

색상
 T1 글자 크기
 14 ▼

▶ 노드 설정

크기 빈도 기준 ▼

- ① **시각화 결과**
클러스터링 분석 결과는 네트워크 차트로 제공됩니다.
Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
Tip. 단어 빈도를 기준으로 설정하면 자주 등장하는 단어가 더 큰 노드로 표시, 매개 중심성을 기준으로 하면 네트워크 내에서 중요한 역할을 하는 단어가 더 큰 노드로 표시됩니다.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
Tip. 마우스 휠 스크롤을 상/하로 움직이면 시각화를 확대/축소해서 볼 수 있습니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

토픽 분석

최적 토픽 개수를 계산하기 위해 파라미터를 조정합니다.

1 분석 알고리즘

최적의 토픽 개수를 판단하기 위한 Perplexity와 Coherence 점수를 계산합니다.
 ※ Perplexity (복잡도) : 모델의 정확도 판단 지표로, 점수가 낮을수록 정확도가 높습니다.
 ※ Coherence (융집도) : 토픽의 의미론적 일관성 판단 지표로, 점수가 높을 수록 일관성이 높습니다.

2 파라미터 선택

- Alpha: 0.5
- Beta: 0.01
- Iterations: 50
- 토픽 수: 10
- 토픽 단어 수: 30
- 랜덤 값: 사용 / 사용안함

3 적용

3 파라미터 선택

- Alpha: 문서의 토픽 분포를 조정하는 매개변수
- Beta: 토픽의 단어 분포를 조정하는 매개변수
- Iterations: 모델 훈련 시 반복 횟수
- 토픽 수: 만들고 싶은 그룹의 개수
- 토픽 단어 수: 그룹 내 들어갈 단어의 개수
- 랜덤 값: 무작위 할당을 진행 (단, 실험의 결과를 재현하거나 검증하는데 어려움)

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

② **파라미터 선택**
LDA 분석에 앞서 최적 토픽 성능을 얻기 위해 파라미터를 기본값으로 설정하거나 사용자 설정으로 선택할 수 있습니다.
 Tip. 기본값을 사용하는 것을 권장합니다. 이후 토픽 분석 결과를 확인한 후 필요에 따라 사용자 설정을 통해 파라미터를 조정하세요.

③ 분석을 적용합니다.

토픽 분석

최적 토픽 성능 계산 결과를 확인합니다.

- <58> -

TEXTOM 데이터 분석 콘솔

단어분석 매트릭스 네트워크 분석 토픽 분석 감성분석 QAP 가설검정

최적 토픽 성능 LDA 토픽모델링

최적 토픽 성능 [분석 알고리즘](#)

최적의 토픽 개수를 판단하기 위한 Perplexity와 Coherence 점수를 계산합니다.
 ※ Perplexity (복잡도) : 모델의 정확도 판단 지표로, 값이 낮을수록 정확도가 높습니다.
 ※ Coherence (응집도) : 토픽의 의미론적 일관성 판단 지표로, 값이 높을수록 일관성이 높습니다.

▶ 파라미터 선택

기본값 사용자설정

▶ Alpha 0.5

▶ Beta 0.01

▶ Iterations 50

▶ 토픽 수 10

▶ 토픽 단어 수 30

▶ 랜덤 값 사용 사용자선택

적용

최적 토픽 결과

성능 계산 결과 시각화 결과 [다운로드](#)

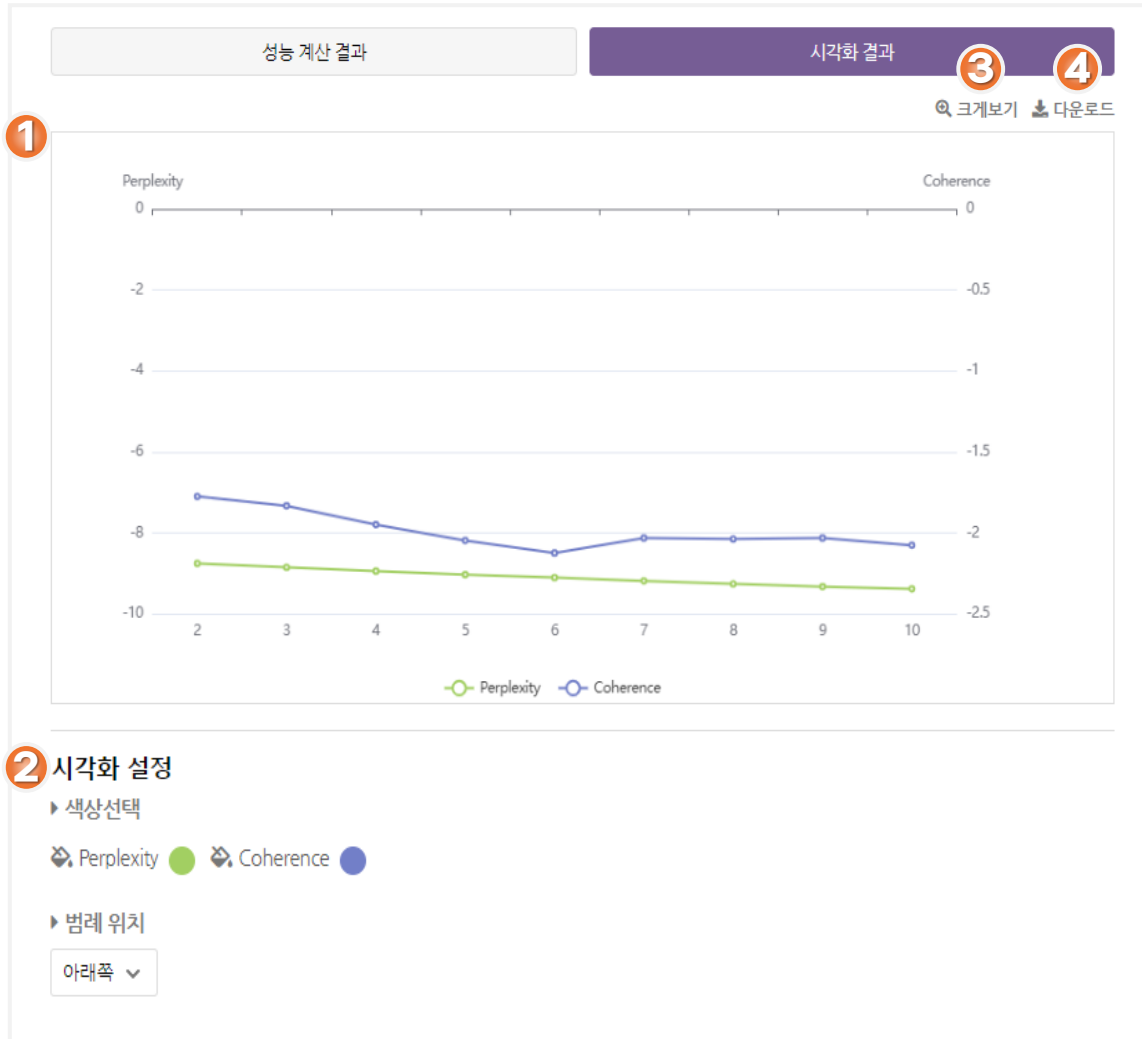
토픽 수	Perplexity Score	Coherence Score
2	-8.758290124036307	-1.7751822255060405
3	-8.854337917772677	-1.8330323423999475
4	-8.94590953066887	-1.9492661466282826
5	-9.032863867924979	-2.0471140930726213
6	-9.105834589867127	-2.1249146381044866
7	-9.186344912080973	-2.032391559779728
8	-9.263125724837977	-2.038861192223454
9	-9.329591527087922	-2.0319589797061535
10	-9.383020447397556	-2.076651981629602

- ① 토픽성능 분석 결과
토픽성능 결과로 다음의 지표를 산출합니다. 일반적으로 Perplexity 값이 낮으면서, Coherence 값이 높을 때, 최적의 토픽 개수를 갖는다고 볼 수 있습니다.

- **Perplexity Score (복잡도)**
토픽모델의 예측 성능 지표로, 값이 작을수록 값이 낮을수록 모델의 예측 정확도가 높음을 의미합니다.
- **Coherence Score (응집도)**
토픽 내 단어들의 의미적 일관성을 평가하는 지표로, 값이 높을수록 더 일관된 토픽을 나타냅니다.

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

- ② 분석 결과 다운로드
군집 및 단어 목록을 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.



- ① **시각화 결과**
최적 토픽 성능 분석 결과는 선그래프로 표현됩니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

토픽 분석

LDA 토픽모델링 결과를 확인합니다.

LDA 토픽모델링 결과

1 토픽 단어 토픽 문서 LDAvis 다운로드

토픽 구분	토픽 포함단어	토픽 포함확률
1	전세사기	0.06
1	전세	0.047
	보증금	0.019
	집	0.018
	수	0.018
	계약	0.017
	반환	0.016
	사기	0.016
	것	0.014
	경우	0.012
	주인	0.012
	소송	0.012

2 토픽 단어 토픽 문서 LDAvis 3 다운로드

원문	토픽 구분	토픽 포함확률
전세사기 신고 전세 보증금 미 반환 피해자 높 전 전세사기 피해 발생 빠른 대응 부동산 투자 과정 형성 방식 선호 전세 계약 사기 전세 보증금 미 반환 금전 피해 일 발생 전세금 경우	1	0.6985715
부동산 전세사기 법률 논점 전세사기 종류 응수 것 사기 사건 임대인 위임 관리자 월세 약정 세입자 측 전세 계약 전착 중간 보증금 일 세 지불	7	0.7458767
전세사기 예방 교육 음악회 금요일 저녁 전세사기 예방 교육 분 퇴근 저녁 산 분 퇴근 하나 버스 분 뒤 똑같 일 주일 만강사 재미 있 것 회차 종료 시간 짧 나중	3	0.56776744
제 전세사기 피해자 의뢰인 씨 집 주인 씨 전세 계약 체결 문제 부동산 등기부 집 소유 탁사 거기 수사 과정 씨 사기 최 처벌 수 것 판단 전세사기 경우 예방	1	0.5273182
전세사기 바지 임대인 문제 상황 사기 사건 신속 대응 먼저 전세사기 바지 임대인 구속 실행 위기 전문가 투자 부동산 경기 좋 때 문제 집값 상태 세입자 때 기존 전세	9	0.61178994
전세 작업 대출 대출 사기 집행유예 대응 실행 처벌 전세사기 작업 대출 연루 든 시기 사람 은행 돈 직업 없 처벌 수 것 염두 전세 작업 대출 대출 사기 실행 처벌 위기 집행유예 등 응방 마련 필요	2	0.5314752
전세 보증금 미 반환 전세 보증보험 사기 현명 대처 방안 전세사기 대비 보증보험 가입 필요 전세사기 대비 전세 보증보험 사기 대비 나라 인종 보증보험 가입 추천 해당 보험 대표 것 한국 주택 금융 공사 주택 도시 보증 공사 서울보증보험 등 있	1	0.36842224

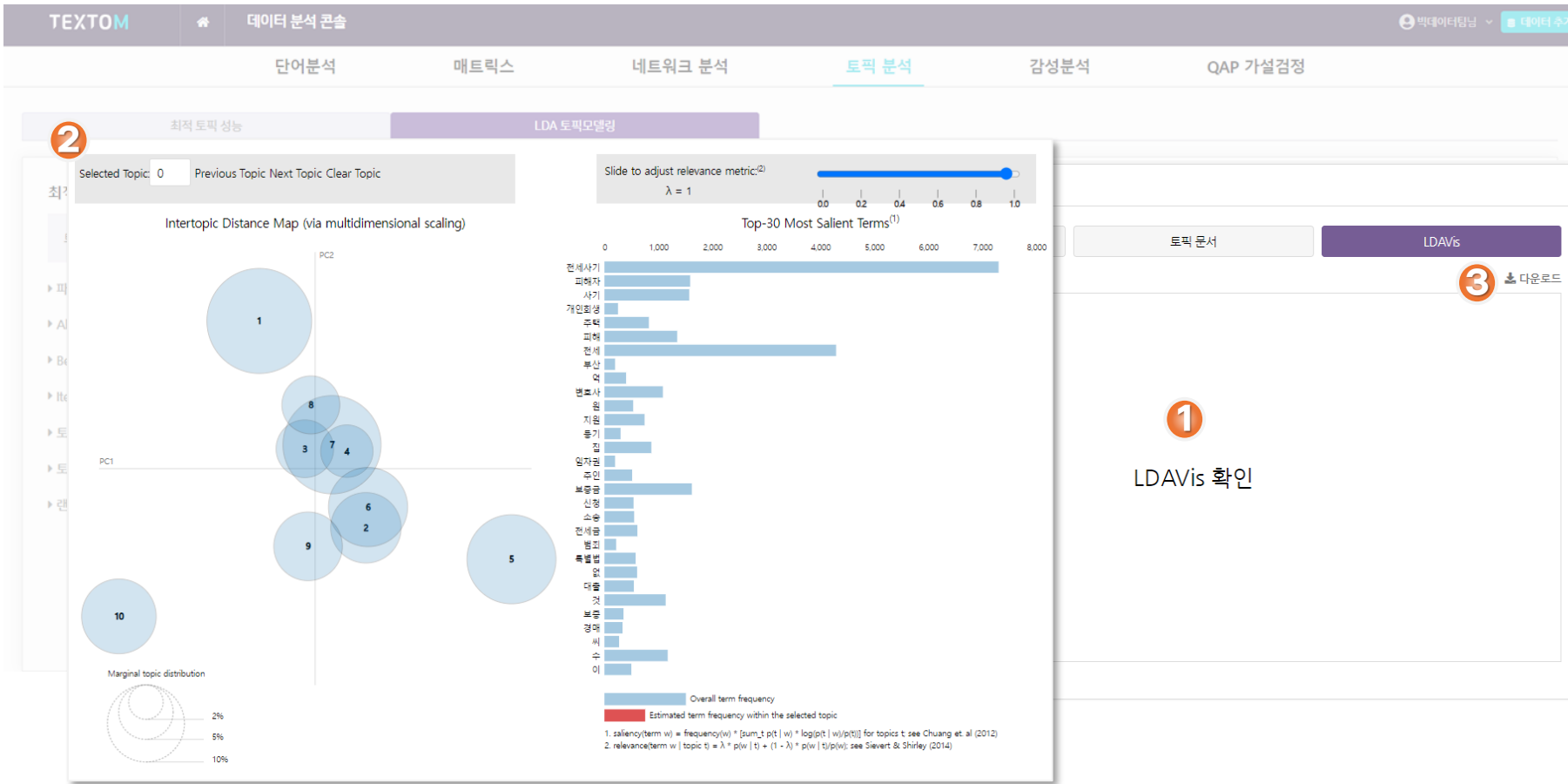
① 토픽 단어 분석 결과
토픽단어 분석결과로 추출된 토픽과 단어의 토픽확률(%)을 산출합니다.
Tip. 단어 토픽확률은 토픽에 해당 단어가 할당될 확률을 의미합니다.

② 시각화 결과
토픽 문서 분석결과로 추출된 토픽에 할당되는 원문과 해당 원문의 토픽확률(%)을 산출합니다.

③ 분석 결과 다운로드
토픽 문서와 토픽 단어 각 분석 결과를 엑셀(.xlsx) 파일로 다운로드할 수 있습니다.

토픽 분석

LDA 토픽모델링 결과를 확인합니다.



- ① [LDAvis 확인] 문구를 클릭하면, 시각화가 팝업으로 나타납니다.
- ② **시각화 결과**
클릭과 커서 이동을 통해 토픽 별 단어 분포와 토픽 간의 유사성 등을 확인할 수 있습니다.
 - 원의 크기: 토픽의 단어 분포를 나타냅니다.
 - 원의 거리: 토픽 간 유사성을 나타냅니다.
 - λ : 토픽 간 차별성을 조절합니다. 1에 가까울수록 한 토픽에서 자주 등장하는 단어가, 0에 가까울수록 다른 토픽에는 자주 등장하지 않는 단어들을 보여줍니다.

Tip. 시각화는 토픽별로 상위 30개의 키워드를 보여주므로, 단어 정제(한 글자, 숫자 등)를 꼼꼼하게 정제하는 것을 권장합니다.
- ③ **시각화 다운로드**
시각화 결과를 html 파일로 다운로드 합니다.

감성 분석

문서분류 감성 분석을 진행합니다.

- <62> -

1 분석 알고리즘

베이지안 분류기(Bayesian Classifier) 기계학습을 통해 텍스트 문서의 감성을 분류하는 방식으로 감성 분석을 진행합니다.

2 예시파일

예시파일을 참고하여 최소 100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링 합니다.
(감성별 학습 데이터의 비율을 비슷하게 업로드 할 수록 정확도가 향상됩니다.)

3 학습데이터

학습데이터 파일을 업로드 해주세요.

업로드한 데이터를 자동으로 분할하여 80%를 학습데이터로, 20%를 테스트데이터로 활용합니다.

4 적용

5 문서분류 데이터를 업로드해주세요.

	A	B
1	비싸서 아무도 안하려고 해요 .개선사례 :	부정
2	Eqc전기자동차는 차량의 구성적인 부분에 긍정	
3	공급대상 차종은 현대 아이오닉과 기아자동차	중립
4	기아자동차국내영업본부장 권혁호 부사장 긍정	
5	안녕하세요전기차 테슬라 모델s 씬팅,블랙	중립
6	자가용발전기등 분산형발전비중을 15%까	중립
7	제가 받은 마그네틱엔 출시를 앞둔전기자동차	중립
8	진정한 친환경은 무한 에너지 태양광을 전	긍정
9	딤싸이클 ★ 선박,전기자동차, 캠핑카에 주	중립
10	기자 =전기자동차(EV)용 전지를 축전용으로	중립
11	전기차가 늘면서 충전소 등 수소 관련 인프라	긍정
12	앞으로전기 자동차가 대세가 되면 부품이	부정
13	케이블 연결 없이 올려놓는 것만으로 거치	중립

- 1** 분석 알고리즘
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- 2** 예시파일
해당 버튼을 클릭하면 예시 형식을 다운로드할 수 있습니다.
- 3** 학습/테스트 데이터 업로드
최소 100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링합니다.
업로드한 데이터를 자동으로 분할하여 80%를 학습데이터로, 20%를 테스트데이터로 활용합니다.
Tip. 감성 별 학습 데이터의 비율을 비슷하게 업로드 할수록 정확도가 향상됩니다.
- 4** 업로드한 데이터 파일을 분석에 적용합니다.
- 5** 분석 결과를 나타냅니다.

TEXTOM
백데이터팀님 | 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

문서분류
감성어휘

문서분류 감성 분석

베이지안 분류기(Bayesian Classifier) 기계학습을 통해 텍스트 문서의 감성을 분류하는 방식으로 감성 분석을 진행합니다.

예시파일을 참고하여 최소 100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링 합니다.
(감성별 학습 데이터의 비율을 비슷하게 업로드 할 수록 정확도가 향상됩니다.)

▶ 학습데이터 🔒

업로드한 데이터를 자동으로 분할하여 80%를 학습데이터로, 20%를 테스트데이터로 활용합니다.

✓ 적용

문서분류 감성 분석 결과

분류 성능
문서 감성 종합
시각화 결과

구분	정확도(%)
전체(Overall Accuracy)	0.0
긍정	0.0
중립	0.0
부정	0.0

[다운로드](#)

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **예시파일**
해당 버튼을 클릭하면 예시 형식을 다운로드할 수 있습니다.
- ③ **학습/테스트 데이터 업로드**
최소 100건에서 최대 1,000건의 데이터를 긍정/중립/부정으로 라벨링합니다. 업로드한 데이터를 자동으로 분할하여 80%를 학습데이터로, 20%를 테스트데이터로 활용합니다.
Tip. 감성 별 학습 데이터의 비율을 비슷하게 업로드 할수록 정확도가 향상됩니다.
- ④ 업로드한 데이터 파일을 분석에 적용합니다.
- ⑤ 분석 결과를 나타냅니다.

분류 성능
문서 감성 종합
시각화 결과

2 [다운로드](#)

1	구분	빈도(건)	비율(%)
	전체(Overall Accuracy)	5914	100
	긍정	4440	75.08
	중립	869	14.69
	부정	605	10.23

3
전체 긍정 중립 부정

4

가장 필요없는 옵션이 선풀루프.

살바엔 뽕겻다

잘봤습니다 차도 예쁘고

안드로이드 오토 블루투스 연결이 안되서 서비스센터가서 모듈 바꿨는데 또 안되네요.. 하 진짜 산거 너무 후회하는중

5 [다운로드](#)

감성분류

긍정

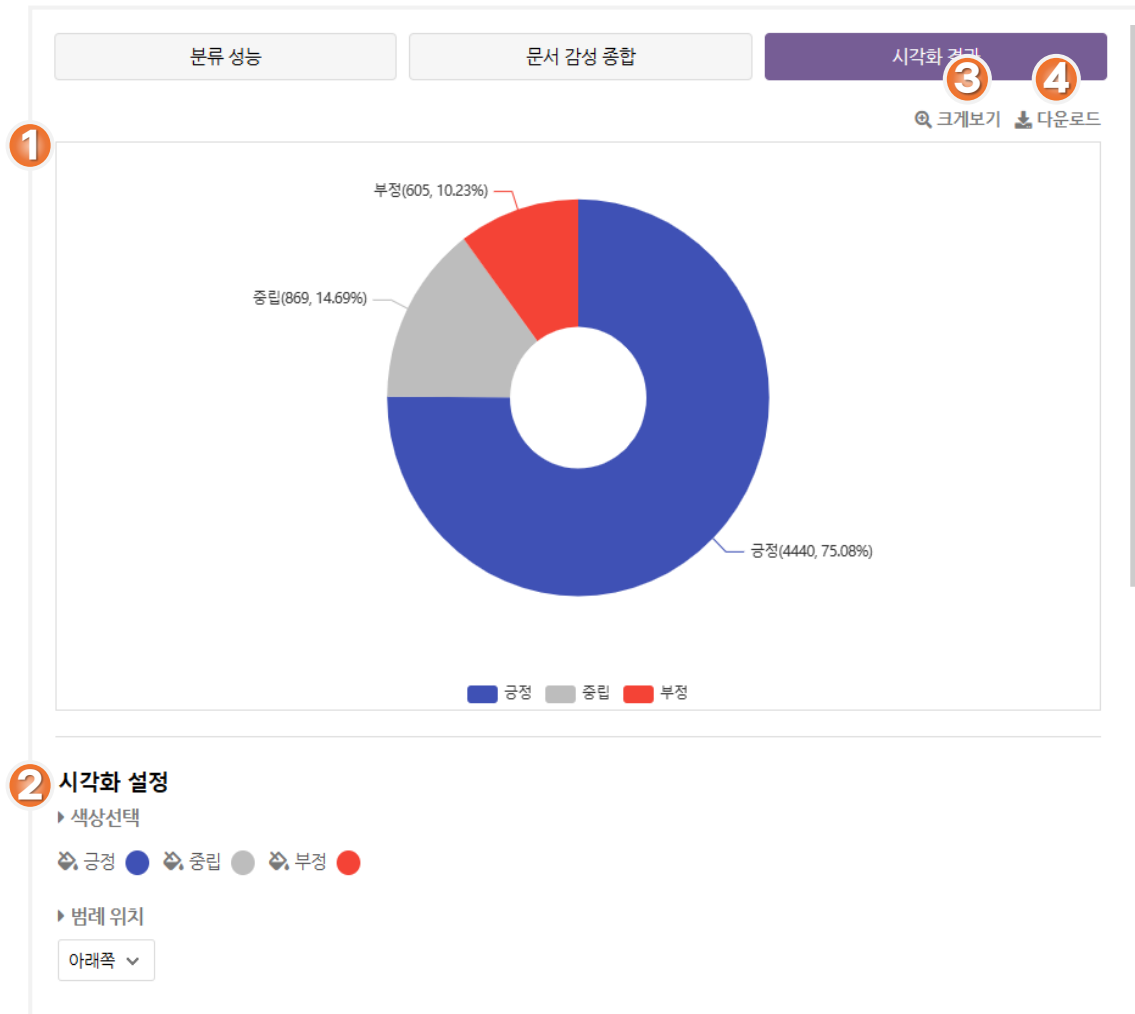
부정

긍정

긍정

부정

- ① **전체 분류 결과 요약**
문서 감성 종합 분석결과로 전체/긍정/중립/부정 문서의 빈도와 비율(%)에 대한 요약 결과를 산출합니다.
- ② **종합 분석 결과 다운로드**
문서 감성 종합 분석 결과를 엑셀(.xlsx) 파일로 다운로드합니다.
- ③ **감성 분류 탭**
감성 탭을 클릭하여 감성 별 원문 분류 결과를 조회할 수 있습니다.
- ④ 감성 별 원문 분류 결과를 확인할 수 있습니다.
- ⑤ **원문 분류 결과 다운로드**
감성별 원문 분류 결과를 엑셀(.xlsx) 파일로 다운로드합니다.



- ① **시각화 결과**
문서 분류 결과는 파이차트로 제공됩니다.
- ② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.
- ③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.
- ④ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

감성 분석

감성 어휘 분석 결과를 확인합니다.

감성어휘 분석 결과

감성어휘 종합
 시각화 결과

1

세부감성 구분	감성분류	빈도(건)	비율(%)
전체		3117	100
호감(good feeling)	긍정	962	30.86
거부감(disgust)	부정	684	21.94
흥미(interest)	긍정	676	21.69
기쁨(joy)	긍정	99	3.18
두려움(fear)	부정	332	10.65
슬픔(sadness)	부정	269	8.63
놀람(fright)	부정	28	0.90
분노(anger)	부정	32	1.03
통증(pain)	부정	35	1.12

2 다운로드

3

전체 호감 거부감 흥미 기쁨 두려움 슬픔 놀람 분노 통증

4

감성어휘	감성강도	빈도(건)	비율(%)
좋다	4.3333	179	5.74
안심	3	116	3.72
안정적이다	3.3333	115	3.69
빠르다	2.7778	90	2.89
신속하다	4.2222	50	1.60
꼼꼼하다	5.1111	50	1.60
추천	4.7778	31	0.99
편하다	5.8889	29	0.93
지적이다	4	24	0.77

5 다운로드

① 세부감성 분석 결과
아래의 9개의 세부감성별 빈도와 비율 결과를 조회할 수 있습니다.

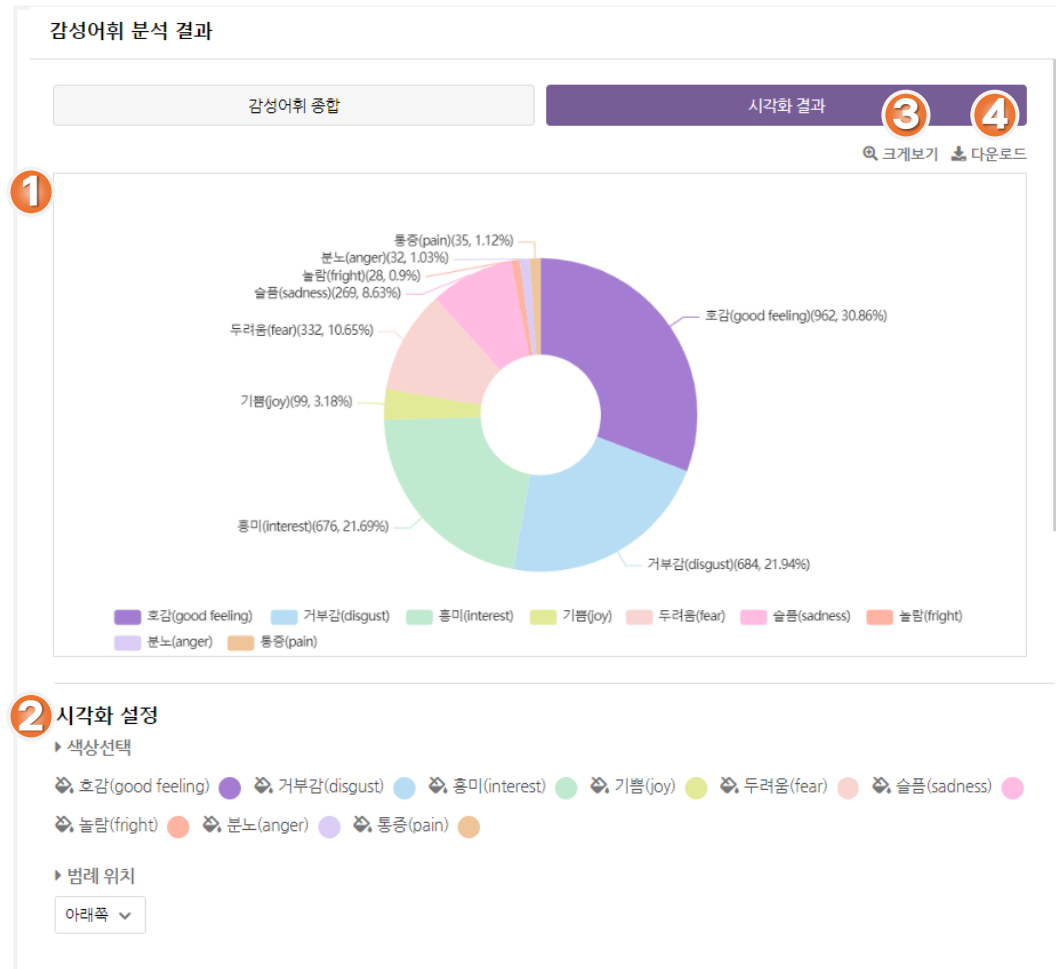
호감	거부감	흥미
기쁨	두려움	슬픔
놀람	분노	통증

② 감성 구분 결과 다운로드
세부감성 분석 결과를 엑셀(.xlsx) 파일로 다운로드 할 수 있습니다.

③ 세부감성 탭
세부 감성 별 탭을 클릭하면, 해당 세부 감성의 감성 어휘를 확인할 수 있습니다.

④ 세부감성 어휘 분석 결과
세부 감성의 어휘, 감성 강도, 빈도, 비율 등을 조회할 수 있습니다.

⑤ 세부감성 어휘 분석 결과 다운로드
세부감성 어휘 분석 결과를 엑셀(.xlsx) 파일로 다운로드합니다.



① **시각화 결과**
감성어휘 분석 결과는 파이차트로 제공됩니다.

② **시각화 설정**
그래프의 다양한 요소를 사용자가 원하는 대로 설정할 수 있습니다.

③ **크게보기**
[크게보기] 클릭 시, 시각화 결과가 팝업창으로 크게 나타납니다.

④ **시각화 다운로드**
시각화 결과를 PNG 파일로 다운로드 합니다.

QAP 분석

QAP 상관분석을 진행합니다.

1 분석 알고리즘

동일한 단어로 구성된 2개 이상의 매트릭스에 대한 상관계수를 바탕으로 네트워크간의 상관성을 분석합니다.
 ※ 피어슨 상관계수 : 두 변수 간의 선형적인 관계를 나타내는 지표로, -1에서 1 사이의 값을 가집니다.
 -1에 가까울수록 음의, 1에 가까울수록 양의 상관관계를 나타냅니다.
 ※ 자카드 계수 : 두 집합 간의 공통된 노드의 비율을 나타내는 지표로, 0에서 1사이의 값을 가집니다.
 1에 가까울수록 두 집합이 유사하다는 것을 의미합니다.
 ※ 유클리디안 계수 : 네트워크에서 두 노드 간의 거리를 나타내는 지표로, 0에 가까울수록 두 점이 가까운 것을 의미합니다.

2 매트릭스

매트릭스 파일을 업로드해주세요. [B]

매트릭스 파일을 업로드해주세요. [B] +

상관관계를 검증할 1-Mode 매트릭스 파일을 2개 이상 업로드해주세요.
 ※ 모든 매트릭스는 동일한 단어로 구성되어야 합니다.

3 파라미터 선택

피어슨 상관계수 자카드 계수 유클리디안 거리

상관성을 측정하기 위한 유사성(상이성) 지표를 선택하세요.

Permutation 3000

4 적용

5 분석할 파일을 업로드해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **매트릭스 파일 업로드**
QAP 분석을 진행할 매트릭스를 업로드합니다. 상관관계를 확인하기 위해 매트릭스를 2개 이상 업로드해주세요.
- ③ **파라미터 선택**
파라미터에서 [유사성 지표]와 [Permutation]을 선택합니다.
Tip. Permutation(순열)은 행렬의 노드들을 무작위로 재배열하여 이를 몇 번 반복할 것인지 정합니다. 통계적 유의성을 평가하기 위해 1,000회 이상 수행하는 것을 권장합니다.
- ④ 업로드한 파일과 조건으로 분석을 진행합니다.
- ⑤ 분석 결과를 나타냅니다.

QAP 분석

QAP 상관분석 결과를 확인합니다.

- <69> -

QAP 상관분석 결과

1

QAP Correlations		QAP P-Values	
	국내_드라마	해외_드라마	다운로드
국내_드라마	1	0.2582237169469755	2
해외_드라마	0.2582237169469755	1	

Correlations (상관계수)

-1.0에서 1.0 사이의 값을 가지며,
0에 가까울수록 두 변수 간의 상관성이 없음을 의미하고,
값이 클수록 두 네트워크가 유사함을 나타냅니다.

P-Value (유의수준)

상관계수가 우연히 나타날 확률을 나타내는 값으로,
낮을수록 상관계수가 통계적으로 유의미함을 나타내며,
일반적으로 P-Values가 0.05 미만일 때 통계적으로
유의미하다고 판단합니다.

3

QAP Correlations		QAP P-Values	
	국내_드라마	해외_드라마	다운로드
국내_드라마	-	2.927564572264821e-17	4
해외_드라마	2.927564572264821e-17	-	

국내와 해외에서의 한국 드라마 선호 패턴을 비교 분석하여 두 시장 간의 유사성과 차이점을 파악하고자 QAP 상관분석을 진행하였습니다.

상관계수 0.2582로 두 네트워크 간에 약한 양의 상관 관계가 있으며,
P-Value 2.927564572264821e-17이라는 극도로 작은 값은 두 네트워크 간의 상관 관계가 우연히 발생할 가능성이 거의 없다는 것을 의미합니다.

- 1 QAP 상관계수(Correlations) 분석 결과
QAP 상관계수 분석 결과를 확인합니다.
Tip. 상관계수는 선형적 관계만을 나타냅니다. 두
변수 간의 관계가 비선형적인 경우
상관계수가 낮게 나타날 수 있습니다.
- 2 상관관계 분석 결과 다운로드
분석 결과를 엑셀(.xlsx) 파일로 다운로드
합니다.
- 3 QAP P-Values 분석 결과
QAP 유의수준 분석 결과를 확인합니다.
Tip. 상관계수가 높고 P-Value가 낮을수록 두
네트워크 간의 상관 관계가 강하고
통계적으로 유의미하다고 해석할 수
있습니다.
- 4 P-Value 분석 결과 다운로드
분석 결과를 엑셀(.xlsx) 파일로 다운로드
합니다.

QAP 분석

QAP 회귀분석을 진행합니다.

- <70> -

TEXTOM
데이터 분석 콘솔
백데이터팀님 | 데이터 추가

단어분석
매트릭스
네트워크 분석
토픽 분석
감성분석
QAP 가설검정

QAP 상관분석
QAP 회귀분석

QAP 회귀분석 1 분석 알고리즘

종속변수 매트릭스에 대한 독립변수 매트릭스의 영향을 평가하는 다중회귀분석입니다. 동일한 단어로 구성된 2개 이상의 매트릭스에 대한 인과관계를 분석합니다.

2 종속변수 매트릭스 업로드

3 독립변수 매트릭스 업로드 +

인과관계를 검증할 1-Mode 매트릭스 파일을 업로드해주세요.
 ※ 종속변수 매트릭스는 1개, 독립변수 매트릭스는 1개 이상 업로드해주세요.
 ※ 모든 매트릭스는 동일한 단어로 구성되어야 합니다.

3 Permutation

4 적용

QAP 회귀분석 결과

QAP Parameters
QAP Regression Coefficients

5 ☰

분석할 파일을 업로드해주세요.

- ① **분석 알고리즘**
버튼을 클릭하면 분석 알고리즘에 대한 설명을 확인할 수 있습니다.
- ② **매트릭스 파일 업로드**
QAP 분석을 진행할 매트릭스를 업로드합니다. 인과관계를 확인하기 위해 1개의 종속변수 매트릭스와 1개 이상의 독립변수 매트릭스를 업로드해주세요.
Tip. 1-Mode 매트릭스여야 하며, 모든 매트릭스는 동일한 단어로 구성되어야 합니다.
- ③ **파라미터 선택**
파라미터에서 [Permutation]을 설정합니다.
Tip. Permutation(순열)은 행렬의 노드들을 무작위로 재배열하여 이를 몇 번 반복할 것인지 정합니다. 통계적 유의성을 평가하기 위해 1,000회 이상 수행하는 것을 권장합니다.
- ④ 업로드한 파일과 조건으로 분석을 진행합니다.
- ⑤ 분석 결과를 나타냅니다.

QAP 분석

QAP 회귀분석 결과를 확인합니다.

QAP 회귀분석 결과

- 1 QAP Parameters
- QAP Regression Coefficients

2 다운로드

구분	값
단어 개수(N)	11
R-Square	0.976831
P-Value	0

QAP Parameters (입력값)

- 단어 개수(N) 변수의 수
- R-Square 결정계수를 나타내며 1에 가까울수록 모델이 데이터를 잘 설명함을 의미
- P-Value 모델의 통계적 유의성을 나타내며 값이 낮을수록 모델의 독립 변수들이 종속 변수에 유의미한 영향을 미친다는 것을 의미

QAP Regression Coefficients (회귀계수)

- 비표준화 계수 독립변수가 종속변수에 미치는 영향
- 표준화 계수 독립변수 간 영향력 비교
- P-Value 모델의 통계적 유의성
- 표준오차 회귀계수의 변동성 및 신뢰성

- QAP Parameters
- 3 QAP Regression Coefficients

4 다운로드

독립변수	종속변수			
	data2			
	비표준화 계수	표준화 계수	P-Value	표준오차
data1	1.494591	0.988348	0	0.022149

- ① QAP Parameters 분석 결과
QAP 회귀분석의 입력값을 분석한 결과를 확인합니다.
- ② QAP Parameters 분석 결과 다운로드
분석 결과를 엑셀(.xlsx) 파일로 다운로드 합니다.
- ③ QAP Regression Coefficients 결과
QAP 회귀계수 분석 결과를 확인합니다.
- ④ QAP Regression Coefficients 분석 결과 다운로드
분석 결과를 엑셀(.xlsx) 파일로 다운로드 합니다.

① **시각화 유형**
 커스터마이징하고 싶은 시각화 유형을 선택합니다.

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

② **시각화 파일 업로드**
 유형에 적합한 데이터 파일을 업로드 하고, [적용] 버튼을 클릭합니다.
 Tip. [예시파일 다운로드]를 활용해 데이터 작성 방법을 확인하세요.

③ **시각화 설정**
 시각화 설정을 통해 시각화 요소의 색상, 크기 등을 변경할 수 있습니다.

④ **시각화 결과 다운로드**
 [다운로드] 버튼을 클릭하여 시각화 결과를 PNG 파일로 다운로드 할 수 있습니다.

The screenshot displays the '마이페이지' (My Page) interface. At the top, there's a navigation bar with '회원정보' (Membership Info), '결제내역' (Payment History), and '사용자사전' (User Directory). The '회원정보' tab is active, showing a table of user details for '빅데이터팀' (Big Data Team). A modal window titled '사용자 인증하기' (Authenticate User) is open, showing an '업로드' (Upload) field and '취소' (Cancel) and '확인' (Confirm) buttons. A red dashed arrow points from the '사용목적' (Usage Purpose) field in the table to the '사용자 인증하기' button in the modal. Below the table, there are buttons for '경부수정' (Modify) and '회원탈퇴' (Withdraw). The '결제내역' tab is also visible, showing a date range filter (2024-05-22 to 2024-05-29) and a '조회' (Search) button. A table of payment records is partially visible below.

아이디	이름	이메일	휴대폰	가입일	용량	사용목적	소속
[Redacted]	빅데이터팀	[Redacted]	[Redacted]	2021-12-14	757.65MB	내부용 (미인증)	[Redacted]

① 회원정보를 조회할 수 있습니다.

② **사용자 인증**
텍스톰 이용 요금 할인을 받으려면 학생증, 사업자등록증, 명함 등 사용 목적을 증빙할 수 있는 서류를 첨부하여 관리자의 승인을 받아야 합니다.
Tip. 학습용 및 R&E용, 교수용 외의 기타 목적의 경우 별도의 인증이 불필요합니다.

③ **정보 수정**
회원정보는 비밀번호 입력 후 수정 가능합니다.

④ **회원탈퇴**
회원탈퇴는 아이디와 비밀번호 입력 후 가능합니다.

⑤ **결제내역**
기간을 설정하여 결제 내역과 관련 정보를 조회할 수 있습니다.

The screenshot shows the '사용자사전' (User Dictionary) management page. It includes a search bar (2), a list of dictionary items (3) with columns for number, change word, correction word, registration date, and usage status. A table with 7 columns is visible. A right-hand panel (1) allows for group selection and file upload. A '변경' (Change) button (4) and a '다운' (Download) button (6) are also present. A '채팅문의' (Chatting) button is at the bottom right.

번호	변경할 단어	수정 단어	등록일	사용여부
168	통종	통종시트	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
167	통종 시트	통종시트	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
166	열선 스티어링	한들열선	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
165	열선 한들	한들열선	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
164	한들 열선	한들열선	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
163	열선 시트	열선시트	2024-04-30 16:54:47	<input checked="" type="checkbox"/>
162	열 시트	열선시트	2024-04-30 16:54:47	<input checked="" type="checkbox"/>

- ① **그룹 선택 및 관리**
그룹을 선택하여 사용자 사전을 관리할 수 있습니다. [수정] 버튼을 눌러 그룹명을 변경할 수 있습니다.
- ② 단어를 검색하여 해당 단어를 포함하는 수정단어 리스트를 확인할 수 있습니다.
Tip. 중복 단어 여부를 확인할 때 사용합니다.
- ③ **파일 업로드**
사용자가 직접 작성한 사용자사전 파일을 업로드하여 [일괄적용] 버튼을 클릭합니다. [예시파일 다운로드]를 활용해 작성 방법을 확인하세요.
- ④ 변경할 단어와 수정 단어를 입력합니다.
- ⑤ 단어 리스트 조회 방식과 단어 사용 여부 등을 선택합니다.
- ⑥ 사용자 사전을 엑셀(.xlsx) 파일로 다운로드합니다.
- ⑦ 체크박스를 선택하여 수정 단어를 삭제할 수 있습니다.

용량 추가

분석에 필요한 데이터 용량을 추가합니다.

TEXTOM 데이터 추가

백데이터팀님 ▼ KOR ▼ 용량 추가

1 데이터 용량/이용기간 선택

선택한 데이터 용량 및 금액: 1MB, 3개월, 4,000원

데이터 용량: 1 MB (추가할 데이터 용량(MB)을 선택해주세요. 데이터 용량은 정수만 구매 가능합니다. (소수점 용량 적용 안 됨.))

이용기간: 3개월, 6개월, 9개월, 12개월 (데이터를 이용할 기간을 선택해주세요.)

2 예산에 맞는 요금제 찾기

금액을 입력하세요. 만원

- ✓ 128MB / 12개월
- ✓ 256MB / 9개월
- ✓ 512MB / 6개월
- ✓ 1024MB / 3개월

결제 예정 금액: 4,000원

상품금액: 4,000원

할인금액: 60만원 이상 할인 5% - 0원

최종 결제금액: 4,000원

3 4,000원 결제하기

구매 이용안내

- 60만원 이상 결제 시 5% 할인 / 150만원 이상 결제 시 10% 할인
- '수집단위'는 10MB 이상 구매 시 사용 가능합니다.
- 문서 1건 당 예상 데이터는 전체수집 약 3KB, 요약수집 약 0.4KB로, 평균적으로 요약수집된 블로그, 카페, 뉴스 1,000건은 300 ~ 500KB입니다. (1MB = 1024KB)
- 논문 1편에 최소 30MB에서 최대 400MB까지 사용되며, 이 데이터량은 연구자가 연구주제에 맞게 설정할 수집 키워드, 기간, 채널에 따라 크게 달라질 수 있습니다.
- 구매한 데이터는 구매 시 선택한 이용 기간까지 사용하실 수 있으며, 이용 기간 만료 시 잔여 데이터는 소멸됩니다.
- 이용 기간 연장 및 소멸 데이터 복구는 불가능하오니, 용량 및 이용 기간을 신중하게 결정하시기 바랍니다.
- 200만원 이상의 결제 및 자체 텍스트 구축은 070-4236-3788 (평일 PM 02:00 ~ PM 05:00), textom@theimc.co.kr 또는 챗봇으로 문의 바랍니다.

데이터 용량 사용안내

- 개인이 보유한 잔여데이터에서 분석에 사용할 용량 만큼 데이터가 차감되는 시스템입니다.
- 데이터 용량은 [데이터수집]단에서 차감이 되지 않습니다.
- 단, 수집완료된 데이터를 분석하기 위해 [정제/형태소분석]단계를 거쳐 데이터분석단의 [텍스트마이닝]으로 넘어갈 때 수집완료된 데이터의 용량만큼 차감됩니다.

구매 데이터 환불정책

- 데이터 구매 후, 1개월 이내에 환불이 가능합니다.
- 단, 구매한 데이터 중 1KB라도 사용을 했을 경우 환불이 되지 않습니다.
- 수집은 무료로 진행할 수 있습니다. 수집된 용량을 확인하시고 그에 맞게 구매 부탁드립니다.

☞ 채팅문의

소개페이지 바로가기

- ① 데이터 용량/이용기간 선택**
 추가할 데이터와 이용 기간을 선택할 수 있습니다.
 TIP. 용량은 1MB 단위별로 구매 가능하며, 3개월/6개월/9개월/12개월 단위로 이용 기간을 선택하실 수 있습니다.
- ② 예산에 맞는 요금제 찾기**
 보유하신 예산에 대한 적정 구매 용량 및 금액을 확인할 수 있습니다.
 TIP. 학교·단체 등 기관 예산 사용이 필요한 경우, 채팅을 통한 문의 부탁드립니다.
- ③ 결제 예정 금액**
 선택한 구매 옵션에 대한 결제 예정 금액이 산출됩니다.
 TIP. [마이페이지] - 회원정보 - 사용목적 에서 소속을 인정할 경우, 목적별 할인 혜택을 제공 받을 수 있습니다.