

# 빅데이터 분석 솔루션 TEXTOM 매뉴얼

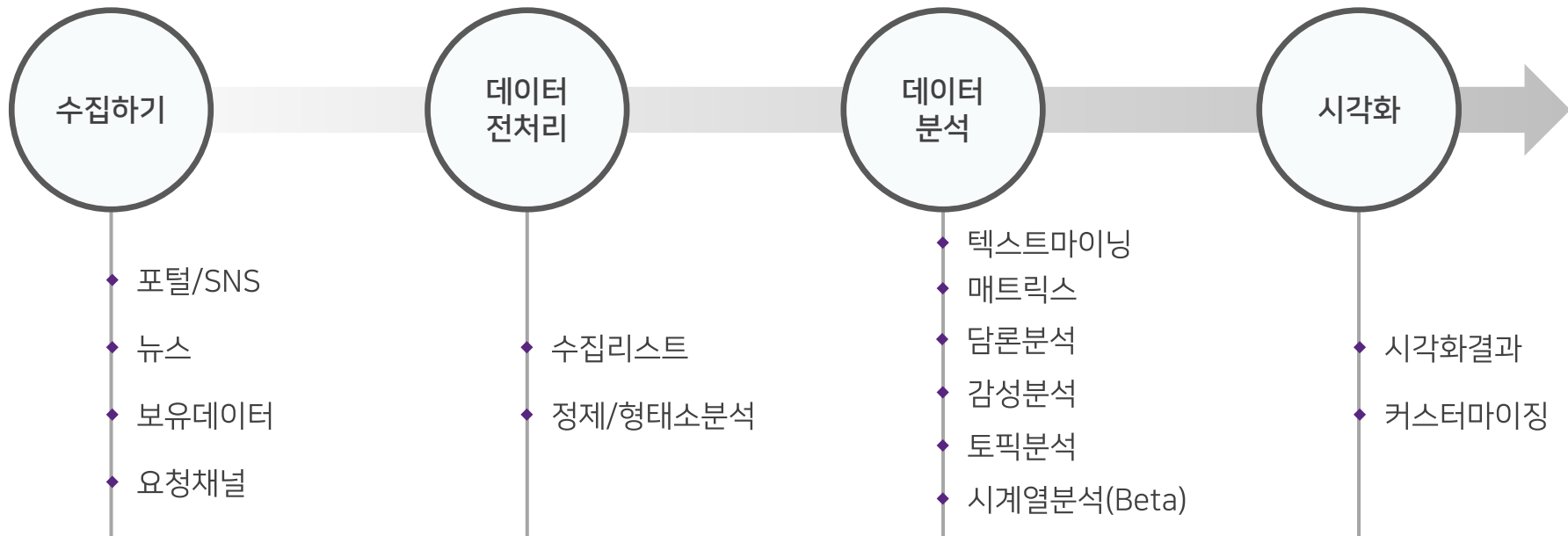
빅데이터 수집에서 시각화까지!

V6.0

TEXTOM

# 한눈에 보는 TEXTOM 프로세스

방대하고 복잡한 텍스트 자료,  
이제 **효율적으로 분석하여**  
연구, 마케팅, 여론분석 등 다양한 곳에 **효과적으로 활용**해보세요!



텍스툼은 Chrome 브라우저에 최적화되어 있습니다

# Contents

1. [회원가입 / 로그인](#)
2. [데이터 수집하기](#)
3. [데이터 전처리하기](#)
4. [데이터 분석하기](#)
5. [데이터 시각화하기](#)

클릭 시 해당 단계 첫 페이지로 이동합니다

# 회원가입 / 로그인

TEXTOM

About

Manual

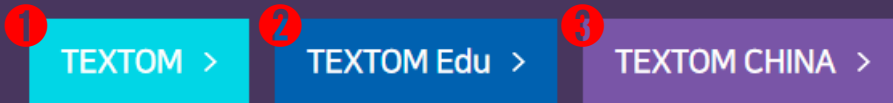
Reference

Communication

Blog

## 빅데이터를 쉽고 빠르게!

분석, 시각화, 인사이트를 다양하게 활용하세요.



수집데이터 706,509건    분석데이터 79,470건    텍스트회원수 18,610명    활용 논문수 309건

ver. 5.5 (2022.02월 현재)

## 원활한 교육을 지원하는 TEXTOM Edu

누구나 교육을 진행 및  
참여할 수 있도록, 출시하였습니다.

OPEN!



< 1 / 2 >

- 1 TEXTOM 데이터 수집, 분석, 시각화까지 쉽고 빠르게, TEXTOM
- 2 TEXTOM Edu TEXTOM으로 교육을 하고 싶은 교수, 강사를 위해 쉽게 교육을 진행할 수 있는 교육용 버전
- 3 TEXTOM CHINA 중국에서 발생하는 데이터를 수집하고, 중문을 분석할 수 있습니다.

※ Edu/CHINA 별도 가입 필요

# 회원가입 / 로그인

**TEXTOM**  
텍스툼을 이용하시려면 로그인이 필요합니다.

아이디  
아이디를 입력해주세요

비밀번호  
비밀번호를 입력하세요

아이디 저장

**로그인**

[아이디/비밀번호 찾기](#) →    [회원가입하기](#) →

**TEXTOM**  
더아이엠씨 님의 접속을 환영합니다!

**START**

[TEXTOM 관리](#) →    [게시판 관리](#) →  
[MY PAGE](#) →    [LOGOUT](#) →

승인 완료 전에는 로그인 후 데이터 수집단계 까지 진행할 수 있습니다.

가입 승인은 영업일 기준 1~3일 이내에 완료됩니다

[회원가입 증명서류 및 무료데이터용량 지급 관련 안내 바로가기](#)

# 로그인 시 첫 화면

데이터수집

포털/SNS   뉴스   보유데이터   요청채널

키워드 미리보기

확인

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

수집키워드

키워드추가

연산자

초기화

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

기간

2022-02-21 ~

2022-02-28

1주

3개월

1년

초기화

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

수집단위

사용

사용안함

채널별로 최대 1,000건의 문서를 수집합니다.

채널

NAVER

네이버 전체

블로그

뉴스

카페

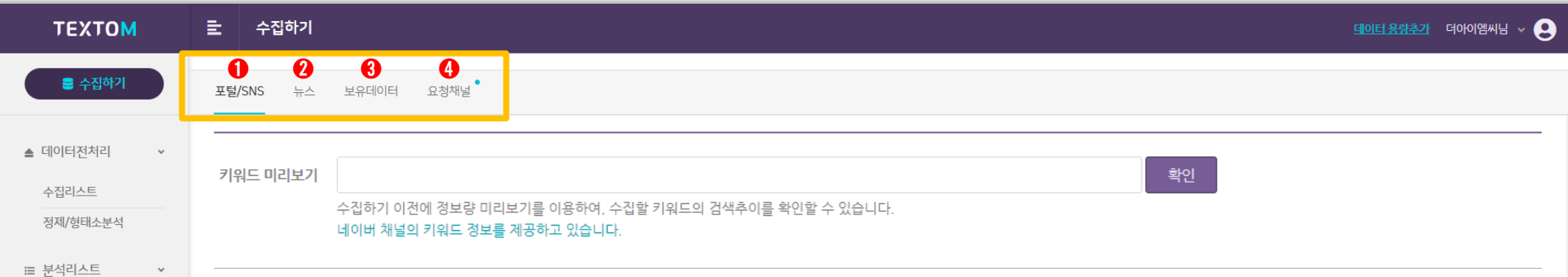
지식IN

학술정보전체

웹문서

채팅문의

수집하기 페이지에서는 분석할 데이터를 수집하거나 가지고 있는 데이터를 업로드할 수 있습니다



- 1 **포털/SNS** 네이버, 다음, 구글, 바이두, 유튜브, 트위터, 페이스북 데이터 수집이 가능합니다
- 2 **뉴스** KBS, MBC, SBS, YTN, 조선일보, 중앙일보, 동아일보, 한겨레, 경향신문 등 언론사 20곳의 기사 데이터 수집이 가능합니다
- 3 **보유데이터** 텍스트로 작성된 pdf, txt, xlsx 형식의 파일 업로드가 가능합니다
- 4 **요청채널** 텍스트롬이 제공하는 채널 외 특정 사이트의 데이터 수집이 필요할 때, 별도의 추가 비용을 지불하고 이용 가능합니다

※ 데이터 수집은 데이터 용량이 차감되지 않습니다

TEXTOM
수집하기
데이터 용량추가 더이엠씨님

수집하기

데이터전처리

수집리스트

정제/형태소분석

포털/SNS    뉴스    보유데이터    요청채널

---

키워드 미리보기  확인

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

## ❶ 수집하기 이전에 정보량 미리보기를 이용하여, 수집할 키워드의 검색추이를 확인할 수 있습니다.

※ 네이버 채널의 키워드 정보를 제공하고 있습니다.

※ 연산자는 적용되지 않습니다.

연관키워드

빅데이터
글로벌빅데이터
통계자료
메가스터디빅데이터
파이썬
청년취업아카데미
무료빅데이터

데이터사이언스
KBS미디어청년취업아카데미
소프트웨어
컴퓨터보안프로그램
앱개발

검색 키워드와 관련도가 높은 키워드입니다.

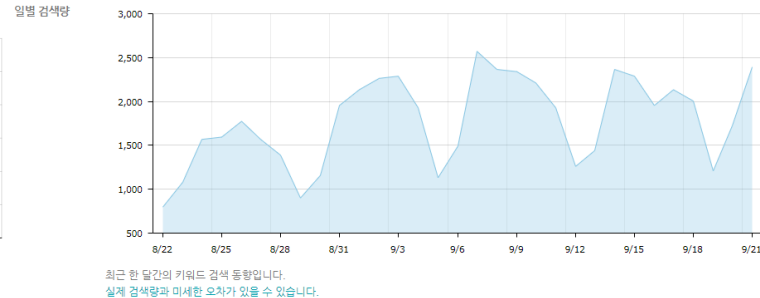
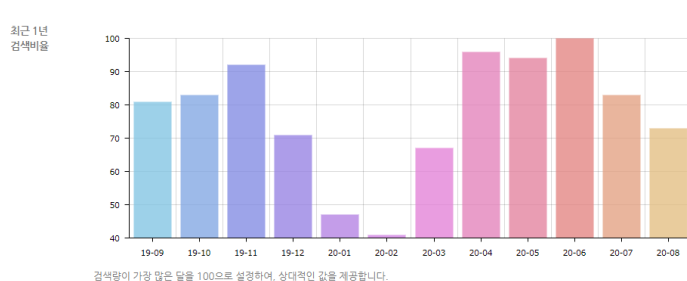
---

월간 검색량

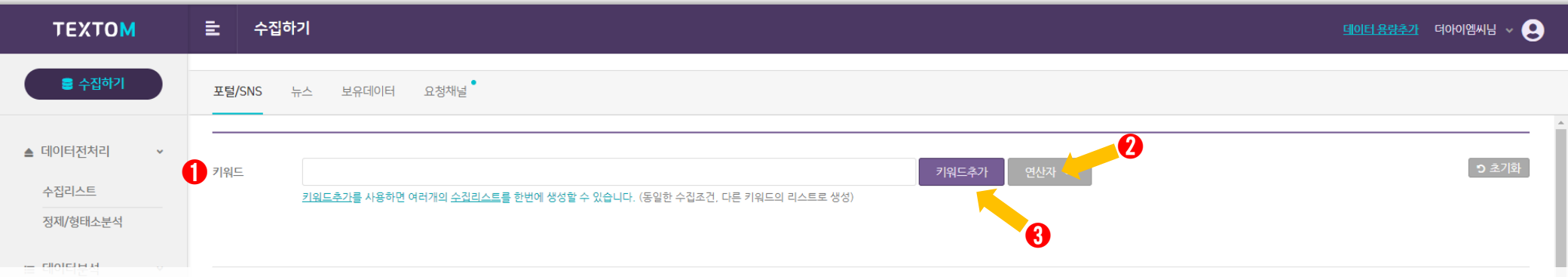
PC
 28,900 건

Mobile
 26,300 건

금일을 제외하고 한달간 키워드가 검색된 횟수입니다.







- ❶ 키워드란에 **기입한 형태 그대로**, 각 채널 검색창에 검색을 하고 그 검색 결과를 수집하게 됩니다
- ❷ 분석의 목적과 주제에 맞는 정확한 데이터를 수집하기 위해서는 세밀한 키워드 선정이 필요합니다  
※ 네이버와 구글은 연산자 기능이 사용 가능한 채널이므로, 상황에 맞게 적절한 특수문자(연산자)를 사용하면 더욱 정확한 데이터를 수집할 수 있습니다
- ❸ **키워드 추가** 기능은 여러 키워드를 동일한 설정(기간, 수집단위, 채널)으로 수집하고 싶을 때 사용하면 유용한 기능입니다.

## ※ 데이터 수집 범위



**[텍스툰으로 분석한 사회 이슈] 주 52시간 근무제** 2018.09.07.  
빅데이터 일관처리 솔루션인 텍스툰으로 짚어보았습니다. ▶ 수집 기간... 도구: 텍스툰(Textom) ▶ 분석 도구: 텍스툰(Textom), UCINet6, NetDraw ▶ 수집데이터량: 41...  
빅데이터 일관처리 ... [blog.naver.com/textom/221354509416](http://blog.naver.com/textom/221354509416)

검색 시 노출되는 게시글의 제목과 요약문을 수집합니다.

# 수집 기간 / 단위

## ❶ 설정한 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 가능하므로, 긴 기간을 수집하려면 요청채널 서비스(유료)를 이용해주시거나 3개월씩 나누어 여러 번 수집해 주셔야 합니다

## ❷ 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 채널은 네이버 블로그, 카페, 지식IN, 뉴스 / 다음 블로그, 카페, 뉴스 / 구글 뉴스입니다

## ① 포털/SNS

채널

수집정보

**NAVER**  네이버 전체  블로그  카페  지식IN  뉴스  웹문서  학술경보전체

**DdM**  다음 전체  블로그  카페  뉴스  웹문서

**Google**  구글 전체  뉴스  구글플러스북  웹문서  
구글플러스북은 구글에서 수집하는 페이스북 문서입니다.

**Baidu**  바이두 **YouTube**  유튜브 **twitter**  트위터 **facebook**  페이스북

## ② 뉴스

채널

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

**KBS**  KBS **MBC**  MBC **SBS**  SBS **YTN**  YTN

**조선일보**  조선일보 **중앙일보**  중앙일보 **동아일보**  동아일보 **한겨레**  한겨레

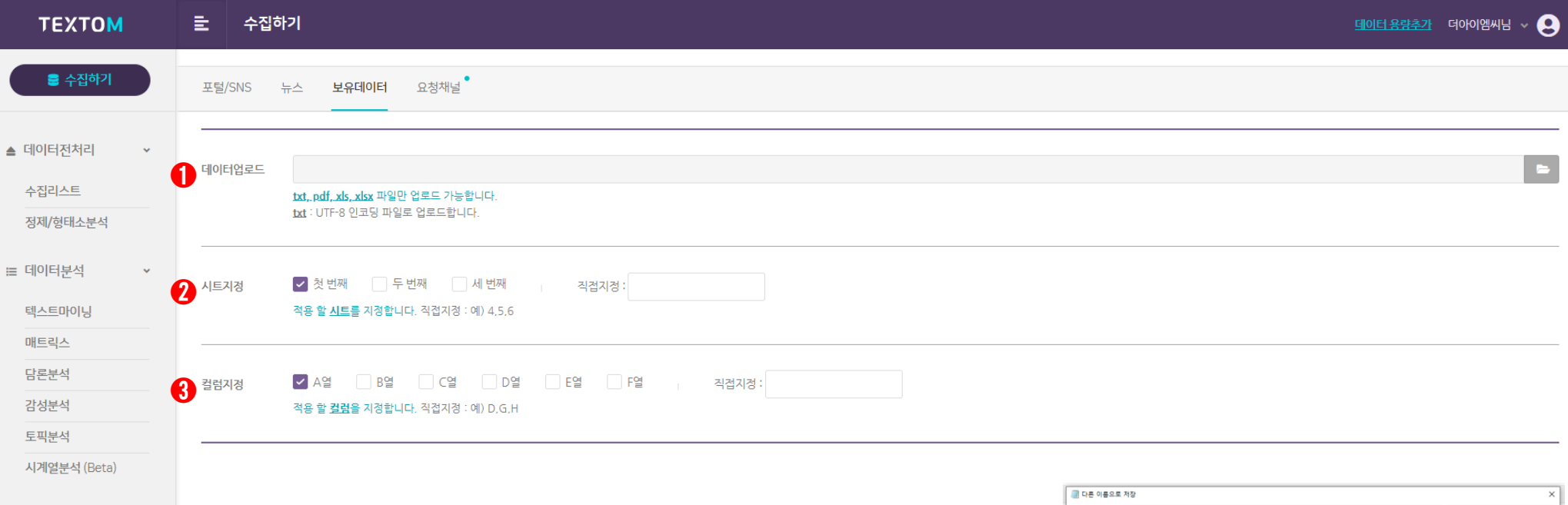
**경향신문**  경향신문 **한국일보**  한국일보 **서울신문**  서울신문 **연합뉴스**  연합뉴스

**news1**  NEWS1 **NEWSIS**  NEWSIS **Oh, my, News!**  Oh,my,News! **노컷뉴스**  노컷뉴스

**매일경제**  매일경제 **한국경제**  한국경제 **전자신문**  전자신문 **ZNet Korea**  ZNet Korea

① 포털/SNS 주요 포털 사이트 및 소셜 네트워크 서비스의 검색 결과를 수집할 수 있습니다

② 뉴스 주요 방송사 및 언론사의 공식 사이트 검색 결과를 수집할 수 있습니다



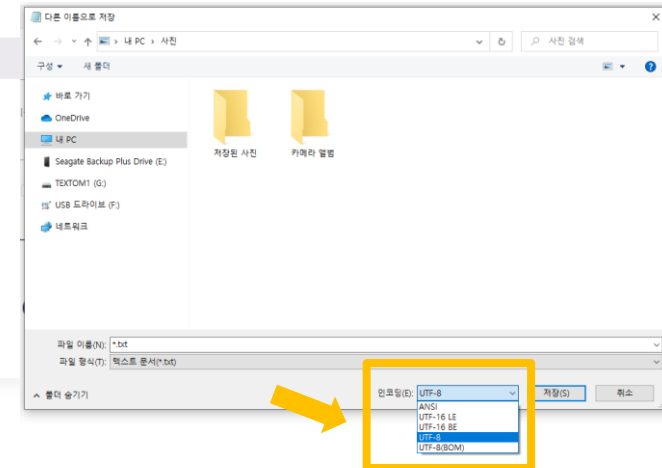
❶ 가지고 있는 데이터를 텍스트롬에 업로드하여 정제, 분석, 시각화 할 수 있습니다

※ 파일 형식은 txt, pdf, xlsx, xls 파일만 가능합니다

※ txt 파일은 UTF-8 인코딩 파일로 업로드 해 주셔야 합니다 [UTF-8 인코딩 설정방법 참고](#) ▶

※ pdf 파일 내 이미지는 인식하지 않습니다

❷, ❸ 시트지정 및 컬럼지정 기능은 엑셀 파일 업로드 시, 많은 시트와 열들 중에서 원하는 데이터만 지정할 수 있는 유용한 기능입니다



※ 컬럼 다중 선택 시, 컬럼들은 한 컬럼으로 합쳐지게 됩니다

❶ 텍스트롬이 제공하는 채널 외의 데이터 수집을 원하면 요청채널로 등록해주세요

※ 적어 주신 내용을 바탕으로 요청 사이트의 데이터 수집가능여부 파악 후, 산출법에 따라 견적이 책정됩니다(추가 유료 서비스)

※ 요청채널은 수집된 데이터만 전달받는 방법(일회성)과 텍스트롬 요청채널 페이지 내 구축하는 방법이 있습니다

※ 수집된 데이터는 xlsx 파일 또는 txt 파일로 전달됩니다

※ 구축이 완료되면 수집하기에서 제공하는 기존 채널과 동일한 방법으로 이용하시면 됩니다

# 수집리스트 살펴보기

TEXTOM
수집완료
데이터용량추기
더아이엠씨님

수집하기

데이터전처리

수집리스트

정제/형태소분석

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

검색결과 15506 / 15506

삭제
정제/형태소분석 →

포털/SNS
뉴스
보유데이터
요청채널

수집완료 리스트의 데이터는 30일이 지나면 삭제됩니다.

<input type="checkbox"/>	키워드	채널	기간	수집날짜	용량	수집상태
<input type="checkbox"/>	강원지방 개발 방안	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체)	2020-03-26 ~ 2020-03-26	2020-03-26	5.00 MB	수집완료
<input type="checkbox"/>	통계/형태소분석	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체), 다음(블로그, 카페, 뉴스, 웹문서), 구글(웹문서, 뉴스, 페이스북) 페이스북 유튜브 트위터 바이두	2020-03-13 ~ 2020-03-20	2020-03-20	8.28 MB	수집완료
<input type="checkbox"/>	코로나 +메르스	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체)	2020-03-13 ~ 2020-03-20	2020-03-20	8.28 MB	수집완료
<input type="checkbox"/>	코로나 +메르스	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체)	2020-03-13 ~ 2020-03-20	2020-03-20	8.28 MB	수집완료
<input type="checkbox"/>	통계/형태소분석	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체), 다음(블로그, 카페, 뉴스, 웹문서), 구글(웹문서, 뉴스, 페이스북) 페이스북 유튜브 트위터 바이두	2020-03-13 ~ 2020-03-20	2020-03-20	8.28 MB	수집완료
<input type="checkbox"/>	통계/형태소분석	네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체)	2020-03-13 ~ 2020-03-20	2020-03-20	8.28 MB	수집완료

### 데이터미리보기

용량을 클릭하면 해당 섹션에서 수집된 데이터 원문을 미리볼 수 있습니다.

▶ 코로나 +메르스  
2020-03-13 ~ 2020-03-20

채널	섹션	수집량(건)	용량
네이버 NAVER N	웹	1,000	437 KB
	블로그	1,000	331 KB
	뉴스	907	377 KB
	카페	1,000	333 KB
	지식인	327	178 KB
	학술정보전체	30	10 KB
다음 D	웹	688	276 KB
	블로그	975	370 KB
	뉴스	772	264 KB
구글 G	카페	680	282 KB
	뉴스	287	124 KB
페이스북 facebook	페이스북	0	0.00
	페이스북	0	0.00

1 수집중에서 수집완료까지 걸리는 시간은 설정해주신 수집 옵션에 따라 최소 10분에서 최대 5일이 소요됩니다

2 데이터미리보기에서 용량의 수치를 클릭하시면 수집된 데이터를 팝업창으로 미리 살펴볼 수 있습니다

※ 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

# 수집단위 사용한 수집리스트 살펴보기

TEXTOM 수집완료

키워드검색: 검색결과 15506 / 15506

수집단위: 10개

수집단위 리스트의 데이터는 30일이 지나면 삭제됩니다.

키워드	채널	기간	수집날짜	용량	수집상태
코로나+메르스	네이버(블로그, 카페, 뉴스, 지식IN) 다음(블로그, 카페, 뉴스) 구글(뉴스)	2020-03-13 ~ 2020-03-20	2020-03-20	6.26 MB	수집완료
2020-03-20 ~ 2020-03-20				213 KB	수집완료
2020-03-19 ~ 2020-03-19				841 KB	수집완료
2020-03-18 ~ 2020-03-18				909 KB	수집완료
2020-03-17 ~ 2020-03-17				965 KB	수집완료
2020-03-16 ~ 2020-03-16				1006 KB	수집완료
2020-03-15 ~ 2020-03-15				688 KB	수집완료

데이터미리보기

용량을 클릭하면 해당 섹션에서 수집된 데이터 원문을 미리볼 수 있습니다.

코로나+메르스  
2020-03-13 ~ 2020-03-20

간별 수집량 시각화

채널	섹션	수집량(건)	용량
네이버 NAVER	블로그	7250	2.39 MB
	뉴스	2042	847 KB
	카페	1678	567 KB
다음 DUM	지식인	307	165 KB
	블로그	3541	1.17 MB
구글 Google	뉴스	2565	866 KB
	카페	608	216 KB
구글 Google	뉴스	310	107 KB

- ① 수집단위 기능을 사용한 데이터는 키워드 앞에 폴더 아이콘이 있으며 해당 리스트를 클릭하면 수집 시 선택한 수집단위로 나뉘어진 하위 리스트가 펼쳐집니다
- ② 나뉘어진 데이터리스트를 개별적으로 선택하시어 데이터미리보기에서 용량의 수치를 클릭하시면 수집된 데이터를 팝업창으로 미리 살펴볼 수 있습니다

※ 수집단위 기능이 뭔가요?  
 일/주/월/년의 시간 단위 중 선택된 단위로 데이터를 최대 1,000건 수집하는 기능입니다  
 Ex. 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 사용한 수집리스트 살펴보기

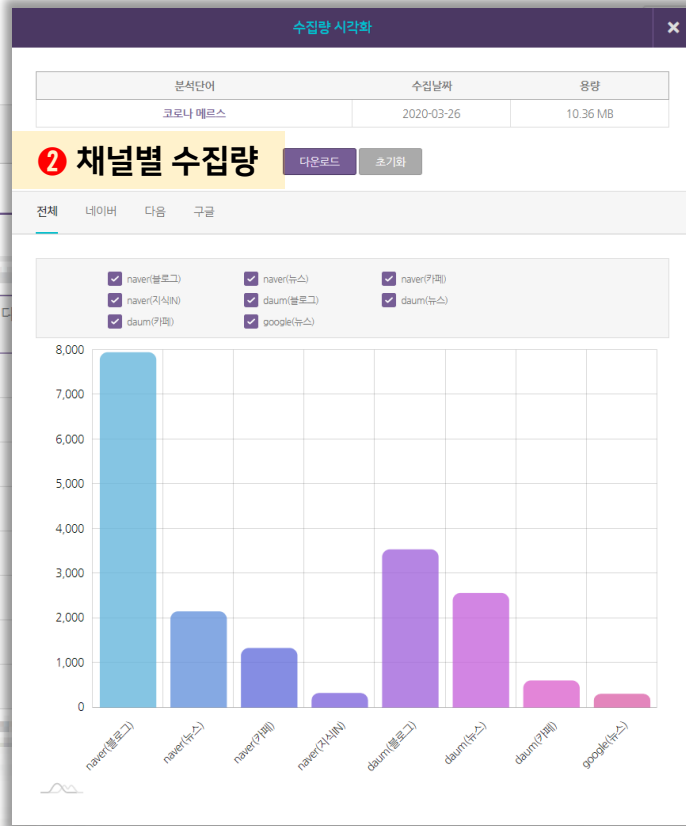
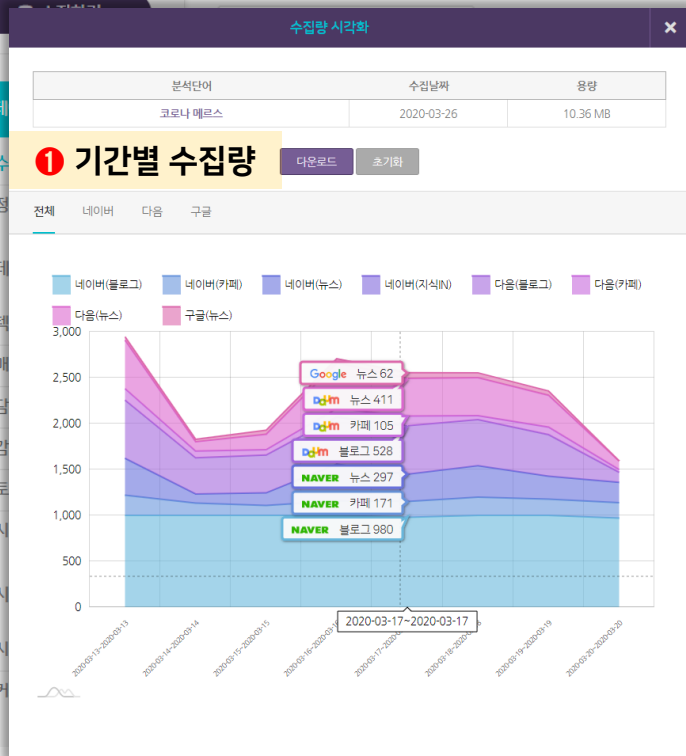
TEXTOM



수집완료

데이터용량추기

더아이멤버님



### 데이터미리보기

용량을 클릭하면 해당 섹션에서 수집된 데이터 원문을 미리볼 수 있습니다.

▶ 코로나 +메르스  
2020-03-13 ~ 2020-03-20

1 기간별 수집량 시각화

채널	섹션	수집량(건)	용량
네이버 NAVER	블로그	7250	2.39 MB
	뉴스	2042	847 KB
	카페	1678	567 KB
	지식인	307	165 KB
다음 Daum	블로그	3541	1.17 MB
	뉴스	2565	866 KB
구글 Google	카페	608	216 KB
	뉴스	310	107 KB

2 채널별 수집량 시각화

정제/형태소분석 →

수집단위 기능을 사용한 데이터는 데이터미리보기에서 기간별, 채널별 수집량 시각화를 확인할 수 있습니다



데이터미리보기

데이터명	생성날짜	용량
코로나+메르스	2020-04-16	7.73 MB

1

네이버
다음
구글

블로그
카페
뉴스
지식인

2

2020.3.13,부동산 경제 뉴스 브리핑  
 realestate.daum.net 강남 3구 하락 이어졌지만 .9억원 미만 서울 외곽 경기 풍선효과 지속 신종코로나바이러스감염증(코로나19) 확산에도 불구하고 풍선효과로...  
<https://blog.naver.com/jayreits?Redirect=Log&logNo=221851157068>

---

2

손 씻기의 중요성, 아무리 강조해도 지나치지 않아요.  
 손 씻기만 잘해도코로나19 같은 많은 질병을 예방할 수 있어요. 제대로 손 씻는... 지남메르스(급성 중동호흡기증후군) 사태 이후, 전국의 의료기관은 물론... m....  
<https://blog.naver.com/ohdamong?Redirect=Log&logNo=221852197601>

---

2

당분간 매매일지는 쉬겠습니다.  
 코로나로 이렇게 망할 줄 몰랐습니다 그저메르스, 사스경도로 생각했는데 참 사는데 막막하겠네요 갑자기 작은 도움도 감사히 받겠습니다.....  
<https://blog.naver.com/the816?Redirect=Log&logNo=221851375750>

---

2

[책거울] 2020.3.12  
 거기다가 사스,메르스, 신종플루,코로나까지 5-6년 주기로 바이러스 공황에 시달리고 있고요. 어느덧 살아온 시간의 절반이 공황도 함께 가고 있는데요. 요즘코로나...  
[https://blog.naver.com/studio\\_smallwave?Redirect=Log&logNo=221851762978](https://blog.naver.com/studio_smallwave?Redirect=Log&logNo=221851762978)

3

1
2
3
4
5
6
7
8
9
10
다음 →
>>

## - 데이터미리보기 팝업창 활용법 -

데이터미리보기 영역 각 섹션의 용량을 클릭하시면 왼쪽 이미지처럼 수집된 데이터를 팝업창으로 미리 살펴볼 수 있습니다

- 1 수집한 데이터는 **채널**과 그 채널의 **섹션**별로 분류되어 나타납니다
- 2 데이터미리보기 팝업창에서는 **제목, 본문 요약, URL**을 제공하며, 각 데이터의 **제목**이나 **URL**을 클릭하시면 해당 데이터의 **실제 페이지**로 이동할 수 있습니다
- 3 데이터미리보기 팝업창으로, 수집된 데이터를 전체적으로 미리 살펴보며 Garbage Data(광고성 글, 원하는 주제와 무관한 글)가 있는지, 수집이 정상적으로 잘 되었는지 체크해줍니다

### ※ Tip

Garbage Data 는 정제/형태소분석 페이지에서 **키워드필터링** 기능으로 제거 가능하므로, 해당 문서들의 공통적인 키워드들을 따로 메모해두시길 바랍니다

# 수집리스트 선택하기

TEXTOM 수집완료 데이터마이닝

수집하기

메르스 검색결과 19 / 18177

삭제 정제/형태소분석 →

포털/SNS 뉴스 보유데이터 요청채널

수집완료 리스트의 데이터는 30일이 지나면 삭제됩니다.

키워드	채널	기간	수집날짜	용량	수집상태	
<input type="checkbox"/>	코로나 + 메르스	네이버(블로그, 카페, 뉴스, 지식IN) 다음(블로그, 카페, 뉴스) 구글(뉴스)	2020-03-13 ~ 2020-03-20	2020-03-26	10.36 MB	수집완료
<input checked="" type="checkbox"/>	2020-03-20 ~ 2020-03-20			700 KB	수집완료	
<input checked="" type="checkbox"/>	2020-03-19 ~ 2020-03-19			1.32 MB	수집완료	
<input checked="" type="checkbox"/>	2020-03-18 ~ 2020-03-18			1.44 MB	수집완료	
<input checked="" type="checkbox"/>	2020-03-17 ~ 2020-03-17			1.48 MB	수집완료	
<input checked="" type="checkbox"/>	2020-03-16 ~ 2020-03-16			1.55 MB	수집완료	
<input checked="" type="checkbox"/>	2020-03-15 ~ 2020-03-15			1.11 MB	수집완료	
<input type="checkbox"/>	2020-03-14 ~ 2020-03-14			1.11 MB	수집완료	
<input checked="" type="checkbox"/>	2020-03-13 ~ 2020-03-13			1.67 MB	수집완료	

데이터미리보기

용량을 클릭하면 해당 섹션에서 수집된 데이터 원문을 미리볼 수 있습니다.

코로나 + 메르스  
2020-03-13 ~ 2020-03-20

기간별 수집량 시각화

채널	섹션	수집량(건)	용량
네이버 NAVER	블로그	7950	5.01 MB
	뉴스	2153	1.7 MB
	카페	1334	1015 KB
다음 DUM	블로그	3541	1.17 MB
	뉴스	2565	866 KB
구글 Google	카페	608	216 KB
	뉴스	310	107 KB

채널별 수집량 시각화

정제/형태소분석 →

❶ 수집리스트에서 분석을 원하는 데이터를 선택합니다

※ 다중 선택이 가능하며, 다중 선택한 데이터들은 정제/형태소분석 단계에서 통합하거나 동일한 설정으로 분석리스트를 생성할 수 있습니다

❷ 한 페이지에서 보여주는 데이터 수를 더 늘릴 수 있습니다

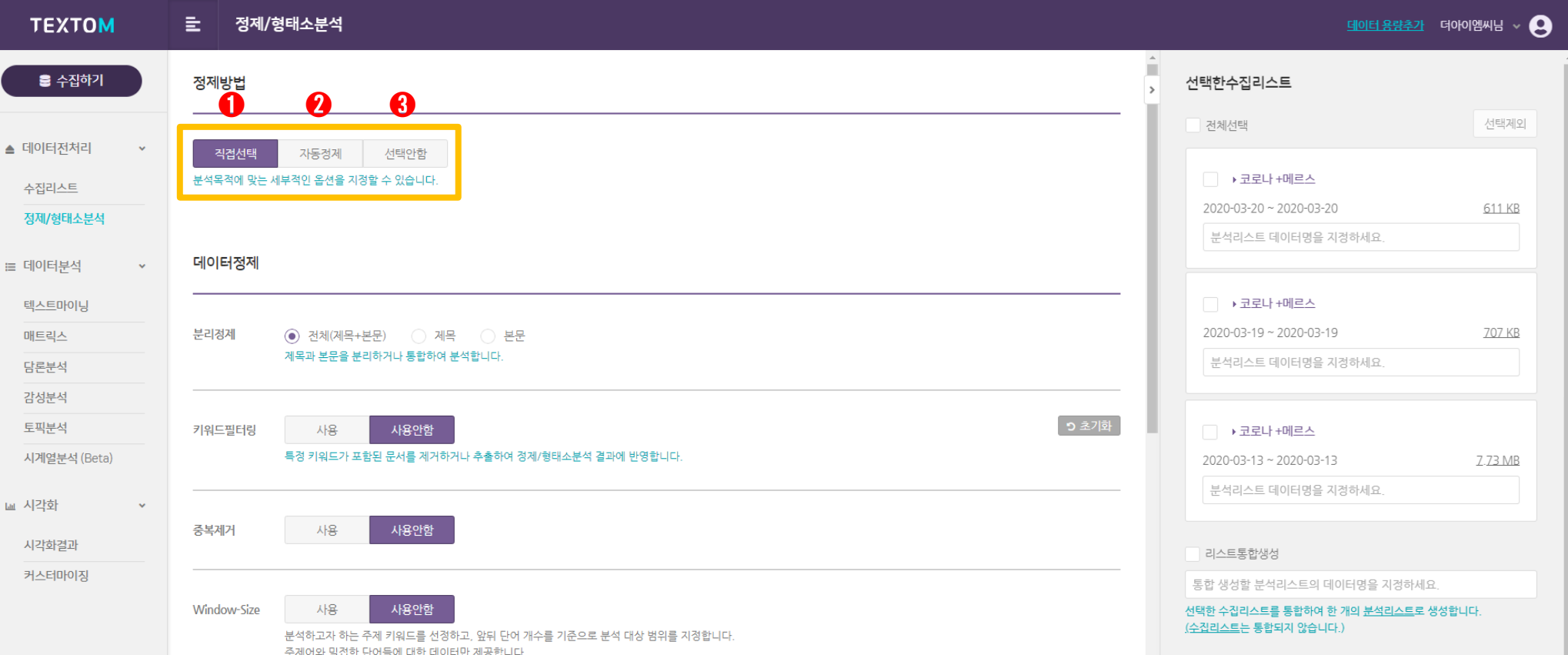
❸ 선택한 데이터가 분석할 데이터가 맞는지 재확인한 뒤, 정제/형태소분석 버튼을 눌러줍니다

# 수집리스트 설정하기

❶ 선택한 리스트에 대한 **데이터명**을 지정하고 체크박스를 선택합니다

※ 데이터명은 생성 시 설정한 분리정제, 분석기, 분석품사 정보를 넣어 지정하시면 추후 여러 데이터 사이에서 구별하기 좋습니다

❷ 리스트를 다중 선택한 경우, 해당 리스트들을 하나의 데이터로 통합하여 생성할 수 있습니다



수집한 데이터 또는 업로드한 보유데이터의 정제, 형태소분석 방법을 지정할 수 있습니다

- ❶ 직접선택 모든 설정을 이용자가 직접 선택할 수 있어 정교한 정제가 가능합니다
- ❷ 자동정제 텍스트롬이 지정해놓은 설정값으로 자동 선택되어, 어느 정도 다듬어진 정제데이터를 볼 수 있습니다
- ❸ 선택안함 이미 정제한 데이터를 업로드하여 분석리스트를 생성하고자 하는 경우 선택안함을 해주시면 됩니다

데이터정제

1 분리정제  전체(제목+본문)  제목  본문  
제목과 본문을 분리하거나 통합하여 분석합니다.

2 키워드필터링   초기화  
 제거  추출  
키워드추가  
특정 키워드가 포함된 문서를 제거하거나 추출하여 정제/형태소분석 결과에 반영합니다.

3 중복제거    
 URL기반  내용기반

4 Window-Size    
키워드:  크기: 3개  
분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상 범위를 지정합니다.  
주제어와 밀접한 단어들에 대한 데이터만 제공합니다.

코로나 +메르스  
2020-03-13 ~ 2020-03-13 2.73 MB  
분석리스트 데이터명을 지정하세요.

리스트통합생성  
통합 생성할 분석리스트의 데이터명을 지정하세요.  
선택한 수집리스트를 통합하여 한 개의 분석리스트로 생성합니다.  
(수집리스트는 통합되지 않습니다.)

수집리스트

분석리스트생성 →

- ❶ 분리정제 수집된 데이터에서 분석에 사용할 텍스트 영역을 지정할 수 있으며, 제목, 본문, 제목+본문 중에서 선택 가능합니다
- ❷ 키워드필터링 특정 키워드가 포함된 문서들을 제거하거나 추출할 수 있습니다

※ 한 필드에 한 단어 씩 기입해 주셔야 하며, 기입해주신 단어들 중 문서 내에 한 단어라도 있으면 적용이 됩니다

※ 제거 : Garbage Data 제거에 유용한 기능으로, 해당 키워드가 속한 문서들을 제거한 정제데이터를 확인할 수 있습니다

※ 추출 : 수집데이터에서 특정 키워드가 포함된 문서를 추출하여 정제할 수 있습니다

# 데이터정제 설정하기

데이터정제

1 분리정제  전체(제목+본문)  제목  본문  
제목과 본문을 분리하거나 통합하여 분석합니다.

2 키워드필터링  사용  사용안함 초기화  
 제거  추출  
키워드추가  
특정 키워드가 포함된 문서를 제거하거나 추출하여 경제/형태소분석 결과에 반영합니다.

3 중복제거  사용  사용안함  
 URL기반  내용기반

4 Window-Size  사용  사용안함  
키워드:  크기: 3개  
분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상 범위를 지정합니다.  
주제어와 밀접한 단어들에 대한 데이터만 제공합니다.

코로나 +메르스  
2020-03-13 ~ 2020-03-13 2.73 MB  
분석리스트 데이터명을 지정하세요.

리스트통합생성  
통합 생성할 분석리스트의 데이터명을 지정하세요.  
선택한 수집리스트를 통합하여 한 개의 분석리스트로 생성합니다.  
(수집리스트는 통합되지 않습니다.)  
 수집리스트

분석리스트생성 →

### 3 중복제거 중복된 데이터를 제거할 수 있습니다

※ URL기반은 링크 전체가 완전히 동일할 때 제거됩니다

※ 내용기반은 특수문자, 띄어쓰기까지 완전히 동일할 때 제거됩니다

### 4 Window-Size 한 문서 내에서 특정 키워드 앞뒤로 단어 개수 범위를 지정할 수 있습니다

**형태소분석**

분석언어:  한국어  영어  중국어

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  품사 태그

▶ 체언  
 일반 명사(NING)  고유명사(NNP)  의존명사(NNB)  단위명사(NNBC)  수사(NR)  대명사(NP)

▶ 용언  
 동사(VV)  형용사(VA)  어근  
 어근(XR)

▶ 수식언  
 관형사(MM)  일반부사(MAG)  접속부사(MAJ)  독립언  
 감탄사(IC)

▶ 접미사  
 명사접미사(XSN)  동사접미사(XSV)  형용사접미사(XSA)  접두사  
 체언접두사(XPN)

▶ 한글 이외  
 외국어(SL)  숫자(SN)

선택한수집리스트

전체선택 선택제외

▶ 코로나 +메르스  
 2020-03-13 ~ 2020-03-20 3.65MB  
 분석리스트 데이터명을 지정하세요.

리스트통합생성  
 통합 생성할 분석리스트의 데이터명을 지정하세요.  
 선택한 수집리스트를 통합하여 한 개의 분석리스트로 생성합니다.  
 (수집리스트는 통합되지 않습니다.)

≡ 수집리스트

분석리스트생성 →

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

❶ 분석언어 데이터를 인식할 언어를 선택해줍니다

❷ 분석기 각 분석기의 특성을 참고하여 형태소분석기를 선택해줍니다 [Espresso K 와 MeCab 분석기 차이 자세히보기](#)

# 형태소분석 설정하기

- 토픽분석
- 시계열분석 (Beta)
- 시각화
- 시각화결과
- 커스터마이징

분석품사

단순품사 ① 상제품사 ④ 품사 태그

초기화 결과 미리보기

**③**

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

▶ 용언  
 동사(VV)  형용사(VA)

▶ 어근  
 어근(XR)

▶ 수식언  
 관형사(MM)  일반부사(MAG)  접속부사(MAJ)

▶ 독립언  
 감탄사(C)

▶ 접미사  
 명사접미사(XSN)  동사접미사(XSV)  형용사접미사(XSA)

▶ 접두사  
 체언접두사(XPN)

▶ 한글 이외  
 외국어(SL)  숫자(SN)

선택한 수식리스트를 통합하여 한 개의 룰리스트로 생성합니다.  
 (수식리스트는 통합되지 않습니다.)

수식리스트

형태소 분석 결과 미리보기 (분석품사)

분석 품사 선택

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

▶ 용언  
 동사(VV)  형용사(VA)

▶ 어근  
 어근(XR)

▶ 수식언  
 관형사(MM)  일반부사(MAG)  접속부사(MAJ)

▶ 독립언  
 감탄사(C)

▶ 접미사  
 명사접미사(XSN)  동사접미사(XSV)  형용사접미사(XSA)

▶ 접두사  
 체언접두사(XPN)

▶ 한글 이외  
 외국어(SL)  숫자(SN)

예시문장	형태소 분석 문장
서울역에서 시침 알라지는 지하철 1호선 한 구간의 거리입니다.	서울역(NG)에서(NG) 시침(NG) 알라(NG)지는(NG) 지하철(NG) 1호선(NG) 한(NG) 구간(NG)의(NG) 거리(NG)입니다(VCP).
이곳의 중간쯤에는 국보 1인 송해문이 있지만 안타깝게도 회마로 불타 지금은 재건축을 위해 가림막을 설치 하든 상태입니다.	이곳(NG)의(NG) 중간(NG)쯤(NG)에는(NG) 국보(NG) 1인(NG) 송해문(NG)이(NG) 있지만(NG) 안타깝(NG)게(NG)도(NG) 회마(NG)로(NG) 불타(NG) 지금은(NG) 재건축(NG)을(NG) 위해(NG) 가림막(NG)을(NG) 설치(NG) 하(NG)든(NG) 상태(NG)입니다(VCP).
그러면 함께 서울역에서 시침 알라지 유용자격하게 걸어 볼까요? LET'S GO!	그러면(NG) 함께(NG) 서울역(NG)에서(NG) 시침(NG) 알라(NG)지(NG) 유용자격(NG)하게(NG) 걸어(NG) 볼(NG)까요(NG)? LET'S GO(NG)! (SF)
짜잔 이국이 서울 4대문 중 하나인 숙정문입니다. 우측을 보시면 치마 옆 쪽에는 소나무의 자태가 곱중스럽습니다.	짜잔(NG) 이(NG)국(NG)이(NG) 서울(NG) 4대문(NG) 중(NG) 하나(NG)인(NG) 숙정문(NG)입니다(VCP)+우측(NG)을(NG) 보(NG)시면(NG) 치마(NG) 옆(NG) 쪽(NG)에는(NG) 소나무(NG)의(NG) 자태(NG)가(NG) 곱중(NG)스럽(NG)습니다(SA)입니다(DF)

선택 품사 적용

## 분석품사

- ① 상제품사 체언, 용언, 어근, 수식언, 독립언, 접미사, 접두사, 외국어, 숫자를 선택할 수 있습니다
- ② 단순품사 명사, 형용사, 동사, 외국어, 숫자를 선택할 수 있습니다
- ③ 결과 미리보기 선택한 품사가 예시문장에 바로 적용되어 정제데이터에 선별될 단어를 미리 확인할 수 있습니다
- ④ 품사 태그 분석기별 품사 태그를 확인할 수 있습니다



# 사용자사전 설정하기

▶ 한글 이외  
 외국어(SL)  숫자(SN)

1 사용자사전 ?    사용    사용안함    2 사용자사전설정

3 그룹지정    미지정

정제할 키워드를 사용자사전에 먼저 등록해주세요. (마이페이지-사용자사전)  
텍스트를 이용하여 여러 번 분석하실 경우, 사용자 사전을 이용하시면 반복적인 작업 없이 빠르고 효율적인 분석이 가능합니다.  
예) 워라밸을 '워라밸'로 변경

- ❶ 사용자사전 유사한 주제의 데이터 또는 동일한 주제의 기간 및 채널만 다른 데이터를 반복 정제해야 하는 경우, 변경할 단어들을 사전으로 구축해두면, 형태소 분석 과정에서 지정해둔 수정 단어들로 일괄 변경되는 기능입니다
- ❷ 사용자사전설정 그룹별로 단어를 등록할 수 있으며, 등록된 단어 사용여부도 매번 설정 가능합니다
- ❸ 그룹지정 그룹지정 란에서 사용할 사전의 그룹명을 선택해 주시면 됩니다

사용자사전

+ 새그룹만들기    삭제

그룹명

미지정

그룹명1

그룹명2

그룹명3

단어검색    검색결과 553671 / 553671    10개

삭제    사용    미사용    다운로드(엑셀)    엑셀일괄등록    전체

변경할단어    →    수정단어    변경

\*'변경할단어'가 아래 리스트에 중복으로 등록된 경우에는 최근에 등록된 내용으로 적용됩니다.  
Tip. 변경할단어의 중복 여부는 상단 '단어검색'을 통해 확인할 수 있습니다.

<input type="checkbox"/>	변경할단어	수정단어	등록일	사용여부
<input type="checkbox"/>	인사	인사정책	2020-05-03 12:09:17	<input type="checkbox"/>
<input type="checkbox"/>	트렌드	트렌드	2020-05-01 15:44:35	<input checked="" type="checkbox"/>
<input type="checkbox"/>	주시꾸뒤르	주시꾸뒤르	2020-05-01 15:44:15	<input checked="" type="checkbox"/>
<input type="checkbox"/>	소매부	소매부분	2020-05-01 02:02:03	<input checked="" type="checkbox"/>
<input type="checkbox"/>	전보	전보다	2020-05-01 02:00:16	<input checked="" type="checkbox"/>
<input type="checkbox"/>	예전	예전에	2020-05-01 01:59:55	<input checked="" type="checkbox"/>
<input type="checkbox"/>	전	전에	2020-05-01 01:58:05	<input checked="" type="checkbox"/>
<input type="checkbox"/>	이번	이번에	2020-05-01 01:57:46	<input checked="" type="checkbox"/>
<input type="checkbox"/>	에슬레저	에슬레저	2020-04-30 18:47:15	<input checked="" type="checkbox"/>
<input type="checkbox"/>	수 있는원피스	수 있는 원피스	2020-04-30 17:08:52	<input checked="" type="checkbox"/>

TEXTOM
텍스트마이닝
데이터용량관리 더아이엠씨님

수집하기

검색결과 23 / 21454

10개

포털/SNS 뉴스 보유데이터 요청채널

	데이터명		생성날짜	용량
<input type="checkbox"/>	코로나 +메르스		2020-03-20	6.26 MB
<input type="checkbox"/>	코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13		2020-03-20	6.26 MB
<input type="checkbox"/>	메르스		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스 -영문		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스 -영문		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스		2020-03-20	1.77 MB
<input type="checkbox"/>	메르스		2020-03-20	1.77 MB

텍스트마이닝
매트릭스
감성분석
토픽분석

형태소 분석이 완료되면, 바로편집하기/업로드를 통해 단어를 경제해보세요. 웹 상에서 빠르고 쉽게 단어 편집을 하고자 할 경우에는 바로편집하기 기능을. 경제 데이터를 내려 받아 작업을 하고자 할 경우 업로드 기능을 사용하세요.

**원문데이터** 1

미리보기
다운로드(Excel)
다운로드(txt)

**정제데이터** 2

미리보기
다운로드(Excel)
다운로드(txt)

**데이터 편집** 3 편집된 데이터가 적용되어 있습니다.

바로편집하기

별도의 다운로드 없이, 웹상에서 데이터를 바로 편집할 수 있습니다

B
✓ 적용

**파일업로드**  
원문데이터가 아닌 정제데이터를 다운로드하여 단어 편집을 진행 후 정제가 완료된 데이터를 업로드 합니다.  
- 엑셀 파일 형식의 데이터를 txt 파일로 변경(UTF-8로 인코딩)하여 단어편집 후 업로드합니다.  
- '편집된 데이터가 적용되어 있습니다'라는 텍스트가 뜨면 파일 업로드 기능을 사용할 수 있습니다.

**1 원문데이터** 수집된 데이터의 원문을 미리보기 혹은 xlsx 파일, txt 파일로 다운로드 할 수 있습니다.

※ 원문데이터의 파일은 데이터 정제 후에도 내용이 변하지 않으며 언제든지 다운로드 받을 수 있습니다.

**2 정제데이터** 정제/형태소 분석 결과 데이터를 미리보기 혹은 xlsx 파일, txt 파일로 다운로드 할 수 있습니다.

※ 정제데이터의 파일은 데이터를 정제 할 때마다 내용이 변경됩니다.

※ 한 번 정제된 데이터는 이전상태로 되돌릴 수 없기 때문에 최초의 정제 데이터를 내려 받아 두고 작업하는 것을 추천해드립니다.

**3 데이터 편집** 고유명사 및 복합명사, 동의어, 불용어 제거 등 데이터 수정을 할 수 있습니다.

**1 바로편집하기** 데이터편집의 한 방법으로, 파일 업로드 없이 웹상에서 바로 데이터를 정제할 수 있는 기능입니다.

**2 정제데이터 미리보기 화면**으로 변경할 데이터 찾기 또는 수정내역이 적용된 데이터를 미리 확인 할 수 있습니다.

**※ 바로편집하기**에서 변경한 내용 적용이 되면 되돌릴 수 없으므로 정제데이터를 내려 받아 두고 작업하는 것을 추천드립니다.

❸ **정확한일치/부분일치** 정확한일치는 정확하게 일치하는 문자열만 변경합니다.

부분일치는 문서에서 부분적으로 일치하는 문자열을 포함하여 모두 변경합니다.

※ 예시) '이번v메르스v사태'와 '이번메르스사태'에서 '메르스'를 '코로나'로 변경하고자 할 때,

- 정확한일치는 띄어쓰기가 있는 '이번v메르스v사태'를 '이번v코로나v사태'로 변경하고, '이번메르스사태'는 변경하지 않습니다.

- 부분일치는 '이번v메르스v사태', '이번메르스사태'를 '이번v코로나v사태', '이번코로나사태'로 '메르스'를 포함한 모든 문자열을 변경합니다.

The screenshot shows the TEXTOM Text Mining interface. The main window is titled '정제데이터 (텍스트마이닝)'. It displays a table of data with columns for '데이터명', '생성날짜', and '용량'. Below the table, there are options for '미리보기' and '바로편집하기'. A search bar and a '변경한 내용 적용' button are also visible. On the right side, a yellow box highlights the '데이터 편집' section, which includes a '바로편집하기' button and a '파일업로드' section. The '데이터 편집' section also shows a '변경' button and a '수정단어' input field.

**④ 단어변경** 좌측 칸에 변경하고자 하는 단어를 입력하고 우측 칸에는 수정된 결과단어를 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.

**⑤ 수정내역** 되돌리기 아이콘을 클릭하면 해당 단어의 변경을 취소할 수 있습니다.

※ 변경한 내용 적용 버튼을 누른 후에는 되돌릴 수 없습니다.

# 텍스트마이닝하기

**6 추천단어** N-gram기반의 확률 모델을 통해 단어쌍을 선별하여 추천합니다.

※ 추천단어 리스트에서 변경하고자 하는 추천단어를 클릭하면 입력을 하지 않아도 바로 수정이 됩니다.

※ N-gram : 데이터에서 2개 단어가 연속적으로 표현된 횟수를 기록한 값

**7 사용자사전 적용하기** 사용자사전 설정 후 이용할 수 있는 기능으로, 적용하기를 누르면 데이터가 수정됩니다.

**8 변경한 내용 적용하기** 수정내역의 내용을 데이터에 적용하고자 한다면 '변경한 내용 적용'버튼을 클릭합니다.

※ 변경 내용을 적용한 후에는 수정 전으로 되돌릴 수 없으니, 수정내역 검토 및 정제데이터를 미리 내려 받아 두는 것을 추천합니다.

The screenshot shows the TEXTOM web application interface. On the left is a navigation menu with options like '수집하기', '데이터전처리', and '데이터분석'. The main area displays a search result for '메르스' with 23/21454 results. Below the search bar is a table with columns for '데이터명', '생성날짜', and '용량'. The table lists several data items, including '코로나 + 메르스' and '코로나 + 메르스 [수집단위] 2020-03-13 ~ 2020-03-13'. On the right side, there is a sidebar with tabs for '텍스트마이닝', '매트릭스', '감성분석', and '토픽분석'. The '텍스트마이닝' tab is active, showing options for '원문데이터' (원문데이터, 다운로드(Excel), 다운로드(txt)), '정제데이터' (정제데이터, 다운로드(Excel), 다운로드(txt)), and '데이터 편집' (데이터 편집, 2 편집된 데이터가 적용되어 있습니다, 바로편집하기). Below this, there are buttons for '파일업로드' (파일업로드) and '적용' (적용).

**❶ 파일업로드** 정제데이터를 다운받아 단어 편집 후, 편집이 완료 된 파일을 업로드 하는 데이터 편집 방법입니다.

※ 원문데이터가 아닌 정제데이터를 다운로드 받아 편집 해야 합니다.

※ 엑셀파일 형식의 데이터를 txt 파일로 저장해야하며, txt 파일은 저장 시 인코딩을 UTF-8로 설정하여 저장 후 업로드해야 합니다.

**❷ 편집한 데이터의 적용 진행 상황**을 확인할 수 있습니다.

※ 데이터 편집이 완료되면 '편집된 데이터가 적용되었습니다.'라는 문구가 등장하고 데이터 확인 후 다시 편집을 진행할 수 있습니다.

※ 단, '편집한 내용을 적용하고 있습니다.'라는 문구가 등장할 경우엔 편집이 진행되고 있는 상황으로 도중에 편집을 진행하면 오류가 생길 수 있습니다.

데이터명	생성날짜	용량
코로나 +메르스	2020-03-20	6.26 MB
코로나 +메르스	2020-03-20	6.26 MB

**1 분석결과 (텍스트마이닝)**

데이터명	생성날짜	용량
코로나 +메르스	2020-05-13	6.26 MB

단어빈도 N-gram TF-IDF 연결중심성

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

단어	빈도	백분율(%)	누적백분율(%)
메르스	296	2.661	2
코로나	277	2.491	5
사스	248	2.23	7
바이러스	238	2.14	9
것	129	1.16	10
19	103	0.926	11
때	92	0.827	12
개발	81	0.728	13
수	73	0.656	13
등	71	0.638	14
신종플루	65	0.584	15
신체	63	0.566	15
신종	60	0.539	16
백신	60	0.539	16
치료제	60	0.539	17

**2 분석결과 (텍스트마이닝)**

데이터명	생성날짜	용량
코로나 +메르스	2020-05-13	6.26 MB

단어빈도 N-gram TF-IDF 연결중심성

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

단어1	단어2	빈도
코로나	바이러스	101
코로나	19	96
사스	메르스	73
신종	코로나	42
메르스	사스	39
메르스	때	34
신종플루	메르스	21
2015년	메르스	19
백신	개발	19
문재인	대통령	19
메르스	사태	18
사스	비교	17
메르스	등	17
치료제	개발	16
바이러스	감염증	15

**파일업로드**  
 원문데이터가 아닌 정제데이터를 다운로드하여 단어 편집을 진행 후 정제가 완료된 데이터를 업로드 합니다.  
 - 엑셀 파일 형식의 데이터를 txt 파일로 변경(UTF-8로 인코딩)하여 단어편집 후 업로드합니다.  
 -편집된 데이터가 적용되어 있습니다라는 텍스트가 뜨면 파일 업로드 기능을 사용할 수 있습니다.

**분석결과**

▶ **단어빈도** 1

미리보기    다운로드(Excel)    다운로드(txt)

▶ **N-gram** 2

미리보기    다운로드(Excel)    다운로드(txt)

▶ **TF-IDF** 3

미리보기    다운로드(Excel)    다운로드(txt)

▶ **개체명인식** 4

미리보기

**1 단어빈도** 추출된 단어와 데이터 내 해당 단어의 빈도수를 나타냅니다.

※ 빈도수가 높다는 것은 정제데이터 내에 해당 단어가 등장하는 빈도가 높다는 것을 의미합니다.

※ 백분율은 전체 언급량을 100으로 했을 때 언급된 양을 의미합니다.

**2 N-gram** n개 단어의 연쇄를 의미합니다.

※ 단어1과 단어2의 빈도가 높다는 것은 두 단어가 나란히 등장하는 빈도가 높다는 것을 의미합니다.



The screenshot shows a web application for text mining. At the top, there are tabs for '데이터명', '생성날짜', and '용량'. Below these, there are two main panels. The left panel, titled '개체명인식' (Named Entity Recognition), shows a list of words and their frequencies. The right panel, titled '분석결과 (텍스트마이닝)' (Analysis Results), shows a table of analysis results for '단어빈도' (Word Frequency), 'N-gram', 'TF-IDF', and '연결중심성' (Link Centrality). The 'TF-IDF' section is highlighted with a yellow box and contains four numbered items: 1. '단어빈도' (Word Frequency), 2. 'N-gram', 3. 'TF-IDF', and 4. '개체명인식' (Named Entity Recognition).

**파일업로드**  
 원문데이터가 아닌 정제데이터를 다운로드하여 단어 편집을 진행 후 정제가 완료된 데이터를 업로드 합니다.  
 - 엑셀 파일 형식의 데이터를 txt 파일로 변경(UTF-8로 인코딩)하여 단어편집 후 업로드합니다.  
 - 편집된 데이터가 적용되어 있습니다라는 텍스트가 뜨면 파일 업로드 기능을 사용할 수 있습니다.

**분석결과**

- ▶ **단어빈도** ①
  - 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)
- ▶ **N-gram** ②
  - 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)
- ▶ **TF-IDF** ③
  - 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)
- ▶ **개체명인식** ④
  - 미리보기

③ **TF-IDF** TF(단어빈도)와 IDF(문서빈도의 역수)를 곱한 값으로 단어가 특정 문서에서 얼마나 중요한 지를 나타냅니다.

※ TF : 문서 내 특정단어의 빈도수 / DF : 여러 문서 내의 특정단어 빈도수 / IDF : DF의 역수

※  $TF - IDF = TF \times 1/DF$

※ 특정 범위 내에서 모든 단어들의 빈도수와 단어가 포함된 문서(정보)들의 빈도수를 구한 후 역수를 취해 곱하여 문서의 중요도를 찾는 방법

④ **개체명인식** 단어를 13개의 개체명 범주에 따라 분류하고 그 빈도수를 나타냅니다.

※ 13개의 개체명 : 사람, 학문, 대상물, 기관, 지역, 문명, 날짜, 시간, 숫자, 사건/사고, 동물, 식물, 금속, 용어

※ 단어빈도, N-gram, TF-IDF, 개체명인식 모두 데이터를 정제를 거칠 때마다 순위 및 내용이 변동됩니다.

# 매트릭스 만들기

TEXTOM 매트릭스 데이터용량추가 더이엠씨님

수집하기

메르스 검색결과 23 / 21556

삭제

10개

데이터전처리

수집리스트

정제/형태소분석

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

커스텀마이징

포털/SNS 뉴스 보유데이터 요청채널

데이터적용 리스트

<input type="checkbox"/>	데이터명	생성날짜	용량
<input type="checkbox"/>	코로나 +메르스	2020-03-20	6.26 MB
<input type="checkbox"/>	코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13	2020-03-20	6.26 MB
<input type="checkbox"/>	메르스	2020-02-24	300 KB
<input type="checkbox"/>	메르스	2020-02-20	2.77 MB
<input type="checkbox"/>	메르스	2020-02-20	2.77 MB
<input type="checkbox"/>	메르스 감염병	2020-02-24	30.05 MB
<input type="checkbox"/>	메르스 질병	2020-02-21	58.95 MB
<input type="checkbox"/>	메르스	2018-12-14	9 KB
<input type="checkbox"/>	메르스	2018-11-29	377 KB
<input type="checkbox"/>	메르스	2018-11-29	5.95 MB

분석단어선택

분석에 활용할 단어를 선택해 보세요. 바로선택하기/업로드를 통해 분석하고 싶은 단어에 대한 결과만을 확인할 수 있습니다.

1-mode 2-mode

바로선택하기

적용

예시파일 다운로드

추출한 단어를 txt 파일/Excel 파일(UTF-8로 인코딩)로 업로드 해주세요.

분석단어

선택단어

미리보기 다운로드

분석결과

단어빈도

미리보기 다운로드

넷지리스트

미리보기 다운로드

정제된 단어들 간의 동시출현빈도 (공출현빈도: co-occurrence)로 작성된 매트릭스 데이터를 제공합니다.  
수학적 분석을 위한 소셜네트워크 데이터의 기본 형태는 테이블 형태의 매트릭스입니다.

※ **공출현** : 전체 텍스트 내(한 행) 특정 범위에서 노드들이 같이 출현하였을 때 이 범위 내에 있는 모든 노드들 간에 의미론적으로 상호 연관되는 관계가 있다고 가정

동시출현 빈도 행렬을 산출하여 분석하고자 하는 해당 데이터의 공출현 빈도, 근접관계 등을 통해 네트워크로 시각화가 가능합니다.  
※ 텍스트롬에서 제공하는 매트릭스를 이용하여 UCINET, NodeXL, Netminer, Gephi 등의 프로그램에서 시각화가 가능합니다.

텍스트마이닝   매트릭스   감성분석   토픽분석

분석단어선택

분석에 활용할 단어를 선택해보세요. 바로선택하기/업로드를 통해 분석하고 싶은 단어에 대한 결과만을 확인할 수 있습니다.

**1** 1-mode   2-mode

단어 선택이 적용 되어있습니다.

**2** 바로선택하기

예시파일 다운로드

추출한 단어를 txt 파일/Excel 파일(UTF-8로 인코딩)로 업로드 해보세요.

**3** 분석단어

선택단어   미리보기   다운로드

**4** 분석결과

단어빈도   미리보기   다운로드

엣지리스트   미리보기   다운로드

유클리디언계수   미리보기   다운로드

코사인계수   미리보기   다운로드

자카드계수   미리보기   다운로드

상관계수   미리보기   다운로드

## 1 분석단어 선택 1-mode 또는 2-mode 분석 방법 선택

※ 1-mode는 단어(메인노드)간의 관계를 분석하고 싶을 때, 2-mode는 단어(메인노드)와 범주(서브노드)간의 관계를 분석하고 싶을 때 선택합니다.

## 2 바로선택하기

분석하고 싶은 단어를 웹상에서 바로 선택하거나, 단어리스트를 직접 업로드 할 수 있습니다.

분석단어선택 (1-mode)

데이터명	생성날짜	용량
코로나+매트릭스	2020-05-13	6.26 MB

단어빈도   TF-IDF

선택단어수   확인   선택단어누락비율   %   다운로드   적용

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

<input type="checkbox"/>	단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/>	매트릭스	296	2.661%	2.661%
<input type="checkbox"/>	코로나	277	2.491%	5.152%
<input type="checkbox"/>	사스	248	2.23%	7.382%
<input type="checkbox"/>	바이러스	238	2.14%	9.522%
<input type="checkbox"/>	것	129	1.16%	10.682%
<input type="checkbox"/>	19	103	0.926%	11.608%
<input type="checkbox"/>	때	92	0.827%	12.435%
<input type="checkbox"/>	개발	81	0.728%	13.163%
<input type="checkbox"/>	수	73	0.656%	13.819%
<input type="checkbox"/>	등	71	0.63%	14.458%
<input type="checkbox"/>	신종플루	65	0.584%	15.042%
<input type="checkbox"/>	사태	63	0.566%	15.609%
<input type="checkbox"/>	산종	60	0.539%	16.148%
<input type="checkbox"/>	백신	60	0.539%	16.688%
<input type="checkbox"/>	치료제	60	0.539%	17.227%
<input type="checkbox"/>	2015년	56	0.504%	17.731%
<input type="checkbox"/>	감염	52	0.468%	18.198%
<input type="checkbox"/>	비교	47	0.423%	18.621%
<input type="checkbox"/>	감염병	47	0.423%	19.043%

## 3 분석단어

선택한 단어와 빈도를 미리보기 또는 다운로드 받을 수 있습니다.

## 4 분석결과

단어빈도와 엣지리스트 그리고 단어 간 공출 현을 통한 유사도 계수를 계산한 방식에 따라 4가지 결과값으로 제공합니다.

※ 작성한 매트릭스로 UCINET를 활용하여 시각화를 하실 수 있습니다.

[UCINET 활용법 알아보기](#)

# 매트릭스 만들기(1-Mode)

분석단어선택 (1-mode)

데이터명	생성날짜	용량
코로나 +메르스	2020-05-13	6.26 MB

1 단어빈도      TF-IDF

2 선택단어수  확인    선택단어누적비율  %    다운로드    적용

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

3

단어	빈도	백분율 (%)	누적비율 (%)
<input type="checkbox"/> 메르스	296	2.661%	2.661%
<input type="checkbox"/> 코로나	277	2.491%	5.152%
<input type="checkbox"/> 사스	248	2.23%	7.382%
<input type="checkbox"/> 바이러스	238	2.14%	9.522%
<input type="checkbox"/> 것	129	1.16%	10.682%
<input type="checkbox"/> 19	103	0.926%	11.608%
<input type="checkbox"/> 때	92	0.827%	12.435%
<input type="checkbox"/> 개발	81	0.728%	13.163%
<input type="checkbox"/> 수	73	0.656%	13.819%
<input type="checkbox"/> 등	71	0.638%	14.458%
<input type="checkbox"/> 신종플루	65	0.584%	15.042%
<input type="checkbox"/> 사태	63	0.566%	15.609%
<input type="checkbox"/> 신종	60	0.539%	16.148%
<input type="checkbox"/> 백신	60	0.539%	16.688%
<input type="checkbox"/> 치료제	60	0.539%	17.227%
<input type="checkbox"/> 2015년	56	0.504%	17.731%
<input type="checkbox"/> 감염	52	0.468%	18.198%
<input type="checkbox"/> 비교	47	0.423%	18.621%
<input type="checkbox"/> 감염병	47	0.423%	19.043%

❶ 단어빈도 또는 TF-IDF 순으로 단어를 볼 수 있습니다.

❷ 선택 단어 수

선택 할 단어의 개수를 입력 후 확인을 누르면 선택 단어누적 비율 이 자동으로 계산되어 나오며 빈도 또는 TF-IDF순으로 선택이 됩니다.

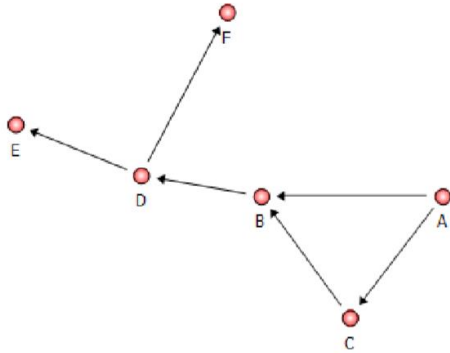
※ 가시성이 좋은 시각화를 만들기 위하여 선택 단어 수를 50~70개 사이로 선택 할 것을 추천 드립니다.

❸ 단어선택

상위 200개의 단어 중 분석하고 싶은 단어를 직접 선택할 수 있습니다.

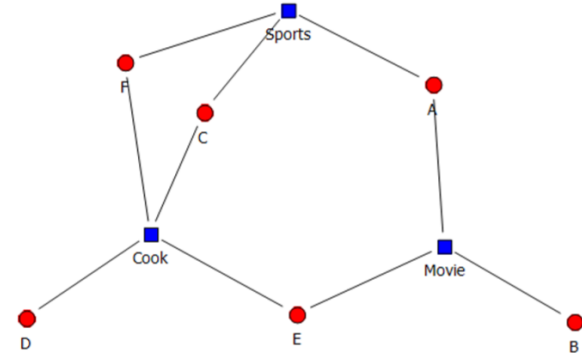
※ 더 많은 데이터를 보고 싶은 경우, 텍스트마이닝 - 분석결과 - 단어빈도 데이터를 다운로드 받아 상위빈도 순으로 전체 단어 리스트에서 원하는 단어를 추출하여 업로드합니다.

## ① 1-mode 분석 방법



	A	B	C	D	E	F
A	0	1	1	0	0	0
B	0	0	0	2	0	0
C	0	3	0	0	0	0
D	0	0	0	0	1	2
E	0	0	0	0	0	0
F	0	0	0	0	0	0

## ② 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

### ※ 1-mode와 2-mode의 차이점은 무엇인가요?

- 1-mode의 경우, 매트릭스와 같이 행과 열이 같은 단어로 이루어진 매트릭스 (모든 네트워크의 가장 일반적인 유형)
- 2-mode의 경우, 매트릭스의 행과 열이 다른 단어로 이루어진 매트릭스 (서로 다른 유형의 조직 간의 관계, 조직과 조직에 속한 조직구성원 간의 관계, 이벤트와 그 이벤트에 참여하는 참석자 간의 관계 등. 예: 기업과 문화재단, 동호회와 회원, 구직자와 취업박람회, 영화와 영화배우 간의 관계)

[2-mode로 분석한 사회 이슈 자세히 보기](#)

# 매트릭스 만들기(1-Mode)

수집하

정제단어 빈도

단어	빈도
메르스	296
코로나	277
사스	248
바이러스	238
것	129
19	103
때	92
개발	81
수	73
등	71
신종플루	65
사태	63
신종	60
백신	60
치료제	60
2015년	56
감염	52

10개

데이터적용 리스트

	생성날짜	용량
	2020-03-20	6.26 MB
	2020-03-20	6.26 MB
	2020-02-24	300 KB
	2020-02-20	2.77 MB
	2020-02-20	2.77 MB
	2020-02-24	30.05 MB
	2020-02-21	30.05 MB
	2018-12-14	9 KB
	2018-11-29	377 KB

텍스트마이닝 매트릭스 감성분석 토픽분석

분석단어선택

분석에 활용할 단어를 선택해보세요. 바로선택하기/업로드를 통해 분석하고 싶은 단어에 대한 결과만을 확인할 수 있습니다.

1-mode 2-mode

1 단어 선택이 적용 되어있습니다.

바로선택하기

적용

예시파일 다운로드

추출한 단어를 txt 파일/Excel 파일(UTF-8로 인코딩)로 업로드 해주세요.

2 분석단어

선택단어 미리보기 다운로드

- 1 단어 선택이 적용 되어있습니다 라는 문구가 나타나면 아래의 분석단어와 분석결과들을 확인 할 수 있습니다
- 2 선택한 단어와 빈도값을 미리보기 또는 다운로드 할 수 있습니다. 다운로드 버튼 클릭 시 엑셀 파일로 다운로드 됩니다.

## 매트릭스결과

- 1 ▶ 매트릭스(빈도)
 

미리보기	다운로드
------	------
- 2 ▶ 엣지리스트
 

미리보기	다운로드
------	------
- 3 ▶ 중심성분석
 

미리보기	다운로드
------	------

### 1 단어빈도의 매트릭스

분석결과 (1-mode) - 빈도

	메르스	코로나	사스	바이러스	것	19	때	개발	수	등	신종	사태	신종	백신	치료	2015	영양	비	성명	국문	화산
메르스	0	207	229	168	83	65	74	36	36	50	51	52	38	40	40	32	32	32	32	32	32
코로나	207	0	176	230	77	147	42	36	40	22	36	44	64	39	39	32	32	32	32	32	32
사스	229	176	0	183	87	45	58	20	23	39	55	17	50	32	32	32	32	32	32	32	32
바이러스	168	230	183	0	71	34	19	35	32	24	18	12	31	21	21	21	21	21	21	21	21
것	83	77	87	71	0	27	30	6	15	8	12	13	7	7	7	7	7	7	7	7	7
19	65	147	45	34	27	0	22	11	25	7	12	16	22	9	9	9	9	9	9	9	9
때	74	42	58	19	30	22	0	0	19	7	31	24	20	12	12	12	12	12	12	12	12
개발	36	36	20	35	6	11	0	0	32	8	12	13	20	12	12	12	12	12	12	12	12
수	36	40	23	32	15	25	19	11	15	7	12	16	20	12	12	12	12	12	12	12	12
등	50	22	39	24	8	7	7	9	19	7	12	16	20	12	12	12	12	12	12	12	12
신종	51	55	36	18	12	31	20	11	25	7	12	16	20	12	12	12	12	12	12	12	12
사태	52	44	17	12	13	16	24	8	19	7	12	16	20	12	12	12	12	12	12	12	12
신종	38	64	50	77	22	7	6	11	25	7	12	16	20	12	12	12	12	12	12	12	12
백신	40	39	32	21	12	9	2	5	19	7	12	16	20	12	12	12	12	12	12	12	12
치료	40	39	32	21	12	9	2	5	19	7	12	16	20	12	12	12	12	12	12	12	12

### 2 엣지리스트의 매트릭스

분석결과 (1-mode) - 엣지리스트

word1	word2	weight
메르스	코로나	207
메르스	메르스	229
메르스	바이러스	168

### 3 중심성분석 계수

중심성분석 계수

단어	연결중심성	근접중심성	매개중심성	위세중심성
온라인	64.977	0.978	0.007	0.293
학습	50.136	0.917	0.004	0.222
전산	56.795	0.936	0.006	0.507
세무	54.205	0.978	0.007	0.494
방식	21.75	0.88	0.004	0.127
관리	28.682	0.978	0.007	0.144
학습	32.0	0.936	0.006	0.304
프로그램	27.182	0.936	0.005	0.15
메뉴얼	8.795	0.423	0.0	0.01
취업	22.841	0.88	0.004	0.152
안내	20.955	0.936	0.006	0.1
뉴스	10.773	0.863	0.004	0.077
서비스	18.523	0.898	0.004	0.107
과목	15.114	0.863	0.004	0.094
공부	9.568	0.83	0.003	0.085
시험	14.25	0.88	0.005	0.112
자격증	19.795	0.83	0.004	0.213
다재됨	15.545	0.863	0.004	0.086

### 1 단어빈도

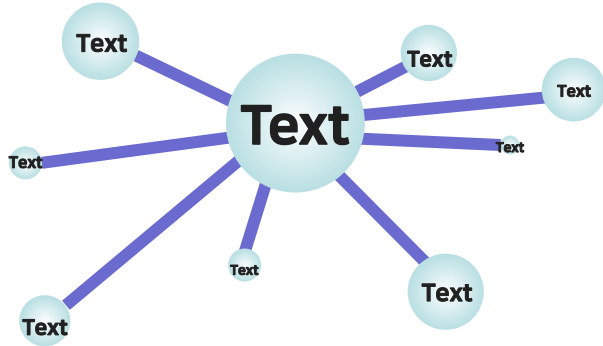
전체 문서 내에서 단어의 출현빈도가 높은 순서대로 단어와 빈도 수를 표시합니다.

### 2 엣지리스트(Edge List)

단어와 단어, 노드와 노드를 짝지어 목록으로 보여줍니다.

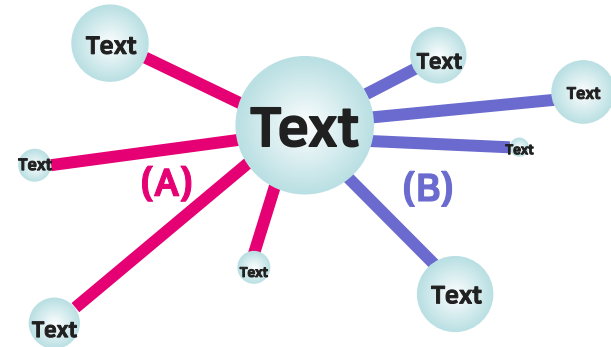
### 3 중심성 분석(Centrality Analysis)

가장 중요한 노드를 찾기 위한 분석입니다.



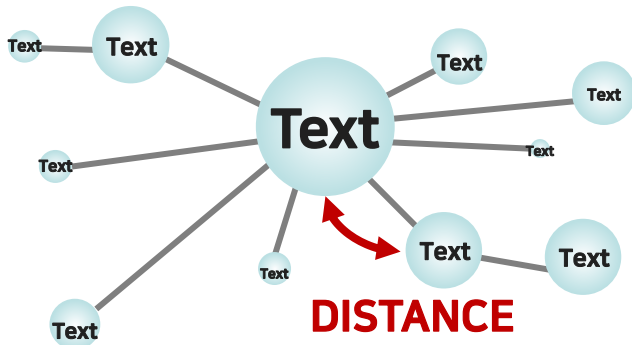
## ① 연결중심성

단어에 다른 단어가 얼마나 많이 연결되어 있는지를 나타냅니다.



## ② 매개중심성

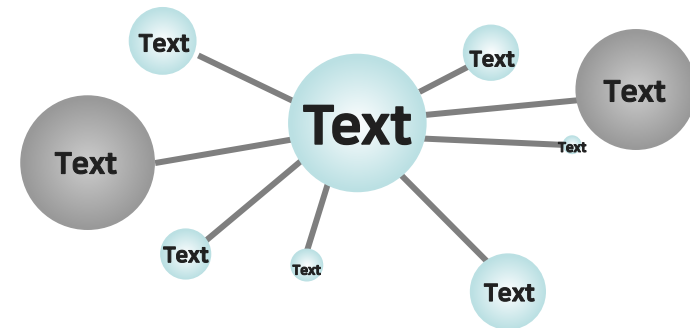
다른 노드와의 연결을 위해 의존성이 높은 노드를 의미합니다.



## ③ 근접중심성

다른 단어와 얼마나 가까이 있는지를 나타냅니다.

노드간 거리 뿐 아니라, 경로의 개수까지 고려합니다.



## ④ 위세중심성

연결된 다른 노드의 중심성을 가중치로 하여 계산한 연결정도입니다.

주변에 중심성이 높은 노드에 연결이 많을 수록 높게 나옵니다.



# 매트릭스 만들기(1-Mode)

**분석결과**

- 단어빈도
  - 미리보기
  - 다운로드
- 엣지리스트
  - 미리보기
  - 다운로드
- 1** 유클리디언계수
  - 미리보기
  - 다운로드
- 2** 코사인계수
  - 미리보기
  - 다운로드
- 3** 자카드계수
  - 미리보기
  - 다운로드
- 4** 상관계수
  - 미리보기
  - 다운로드



분석결과 (1-mode) - 유클리디언계수

분석결과 (1-mode) - 코사인계수

분석결과 (1-mode) - 자카드계수

분석결과 (1-mode) - 상관계수

	메르스	코로나	사스	바이러스	것	19
메르스	0	0.206616444608	0.427450639988	0.127851523891	0.061208212471	0.0464178492521
코로나	0.206616444608	0	0.181898143575	0.329296760968	0.0447057003414	0.512883816455
사스	0.427450639988	0.181898143575	0	0.263346445198	0.152636607957	0.0232936769613
바이러스	0.127851523891	0.329296760968	0.263346445198	0	0.0578307767812	0.0647122687901
것	0.061208212471	0.0447057003414	0.152636607957	0.0578307767812	0	0.0139968784732
19	0.0464178492521	0.512883816455	0.0232936769613	0.0647122687901	0.0139968784732	0
때	0.151405477366	0.0368247070502	0.0987442733342	0.119976786269	0.0618035077386	0.0371320778
개발	0.0577906810547	0.0336550363362	0.108759776196	0.00774767687781	0.102022271365	0.0401109731
수	0.0392215477494	0.0100786613419	0.0873963839952	0.00619315656079	0.0273456851653	0.113078228:
등	0.0849491469634	0.109104771272	0.0463139354076	0.0534878608712	0.0971070225114	0.0798596967
신종	0.118867016678	0.143516107607	0.0437422418514	0.0732097652381	0.0396286543191	0.206026125:
사태	0.130549358618	0.0715859381448	0.099353276482	0.101367827812	0.0235868679661	0.0395405557
신종	0.0396002615727	0.253971550584	0.201218929616	0.362804929854	0.0830940385188	0.0607407984
백신	0.0410663805253	0.0408732533669	0.0242277040026	0.0315193387276	0.0218123415377	0.024767722:
치료	0.0237937753954	0.133141305234	0.060424094882	0.0449446861563	0.0333265118182	0.0880147805
2015	0.106687845169	0.161236803114	0.132107451181	0.0177024890188	0.117436999095	0.0705332533
감염	0.0356642493662	0.0113356670685	0.0125840441496	0.165905946101	0.0485175639136	0.0364299124

- 1** 유클리디언 계수
- 2** 코사인 계수
- 3** 자카드 계수
- 4** 상관 계수

[단어 간의 유사도 측정방법 자세히 알아보기](#)

# 매트릭스 만들기 (2-Mode)

**분석단어선택 (2-mode) 열 단어**

데이터명	생성날짜	용량
2020제주맛집	2022-02-28	479 KB

단어빈도 TF-IDF

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"/> 제주맛집	1198	9.417%	9.417%
<input type="checkbox"/> 제주	844	6.634%	16.051%
<input type="checkbox"/> 맛집	571	4.488%	20.539%
<input type="checkbox"/> 제주도	331	2.602%	23.141%
<input type="checkbox"/> 추천	305	2.397%	25.538%
<input type="checkbox"/> 여행	256	2.012%	27.551%
<input type="checkbox"/> 식당	175	1.376%	28.926%
<input type="checkbox"/> 카페	162	1.273%	30.2%
<input type="checkbox"/> 소개	109	0.857%	31.056%
<input type="checkbox"/> 서귀포	87	0.684%	31.74%
<input type="checkbox"/> 애월	77	0.605%	32.346%

**매트릭스 단어선택**

10개 >

1-mode 2-mode

단어 선택이 적용 되어있습니다.

열 단어 바로선택하기 열

열 단어 선택 완료됐습니다.

행 단어 바로선택하기 행

행 단어 선택 완료됐습니다.

적용

서로 다른 값이 입력된 열과 행 파일을 각각 업로드합니다.  
추출한 단어를 txt 파일/Excel 파일(UTF-8로 인코딩)로 업로드 해주세요.

❶ 텍스트마이닝 데이터를 기반으로 텍스트 파일에 행과 열 각각의 단어를 나열합니다.

(단어 간의 관계를 보여주는 1-mode와 달리 카테고리화 시키는 2-mode는 행과 열에 각각 속성이 다른 단어가 들어갑니다.)

❷ 행/열에 들어갈 단어를 선택하고 적용을 클릭합니다.

※ 열단어를 선택 후 행단어를 선택합니다.

# 매트릭스 만들기 (2-Mode)

**분석단어**

▶ 선택 열단어 미리보기    다운로드

▶ 선택 행단어 미리보기    다운로드

---

**분석결과**

1. 단어빈도 미리보기    다운로드

2. 엣지리스트 미리보기    다운로드

---

▶ 유클리디언계수 미리보기    다운로드

▶ 코사인계수 미리보기    다운로드

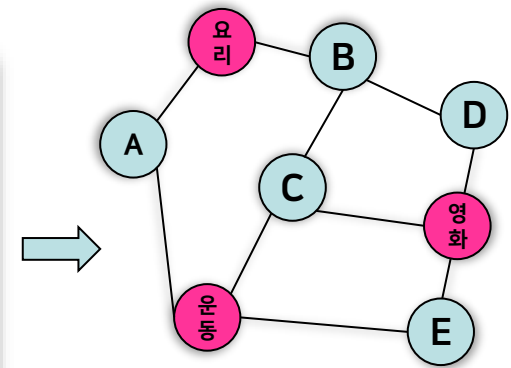
▶ 자카드계수 미리보기    다운로드

▶ 상관계수 미리보기    다운로드

## 1 단어빈도의 매트릭스

분석결과 (2-mode) - 빈도

	개	사	2015	한	전	과	대	인	3	수	대	20	2003	12	우	유	에	영			
메	38	55	0	22	0	22	12	0	0	10	7	14	4	0	0	0	2	17	12	11	5
르	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
스	21	19	0	13	0	28	14	0	0	4	9	15	10	0	0	0	2	17	21	0	4
트	38	12	0	14	0	12	3	0	0	9	5	2	5	0	0	0	1	18	14	2	2
스	18	8	0	7	0	9	1	0	0	4	3	4	3	0	0	0	4	0	1	0	2
스	14	15	0	0	0	2	0	0	0	1	2	3	0	0	0	0	0	0	0	0	2
스	6	6	0	10	0	2	0	0	0	3	0	0	2	0	0	0	0	0	0	0	0
스	58	1	0	0	0	3	0	0	0	2	2	1	0	0	0	0	0	0	0	0	0
스	45	6	0	0	0	0	1	0	0	4	0	1	0	0	0	0	0	0	0	0	0



<Mini Social graph>

## 2 엣지리스트의 매트릭스

분석결과 (2-mode) - 엣지리스트

word1	word2	weight
메르스	개별	38
메르스	사태	55
메르스	2015년	0
메르스	한국	22
메르스	전세계	0
메르스	과거	22
메르스	대유행	12
메르스	문자인	0
메르스	3월	0
메르스	교수	10
메르스	상황	7
메르스	글로벌	14
메르스	중국	4
메르스	2020년	0
메르스	2003년	0
메르스	12월	0
메르스	우리나라	0
메르스	유행	2
메르스	백위	17

### 1 단어빈도

전체 문서 내에서 단어의 출현빈도가 높은 순서대로 단어와 빈도 수를 표시합니다.

### 2 엣지리스트(Edge List)

단어와 단어, 노드와 노드를 짝지어 목록으로 보여줍니다.

TEXTOM

데이터용량추가 더이엠씨님

수집하기

데이터전처리

수집리스트

정제/형태소분석

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

커스터마이징

정제단어 빈도

단어	빈도
교육	133
대학원	119
진학	72
한양대학교	72
기회	69
온라인	68
협약	68
철년	48
전공	48
전문가	41
취득	32
협력	31
성경	25
코로나	25
협력관계	24
안테나	24
교류	22
이상	22

파일명	생성날짜	용량
1	2021-03-01	1.96 MB
2	2021-03-01	1.85 MB
3	2021-03-01	1.27 MB
4	2021-03-01	1.4 MB
5	2021-03-01	1.56 MB
6	2021-02-28	13 KB
7	2021-02-27	3.17 MB
8	2021-02-26	3.11 MB
9	2021-02-26	3.8 MB

10개

문서 내에서 동시에 등장(공출현)하는 단어 사이의 관계를 나타내는 분석으로, 상관관계를 이용하여 단어 간의 관계 패턴에 따라 군집화하는 분석방법입니다.

1 담론 개수

2개	4개
8개	16개

적용

2 분석단어

선택 단어(1-mode)	미리보기	다운로드
선택 열단어(2-mode)	미리보기	다운로드
선택 행단어(2-mode)	미리보기	다운로드

3 분석결과

결과(1-mode)	미리보기	다운로드
밀도값(1-mode)	미리보기	다운로드
결과(2-mode)	미리보기	다운로드
밀도값(2-mode)	미리보기	다운로드

- 1 블록 개수** 담론분석은 각 키워드들의 유사성을 측정하기 위해 상관분석을 반복적으로 실시해 상관관계수 값에 따라 그룹화되는 것을 보며 블록개수를 설정 및 반복하여 최적의 블록개수를 찾습니다.
- 2 분석단어** 매트릭스 단계에서 추출한 단어 미리보기입니다.
- 3 분석결과** 상관분석에서 도출되는 상관관계수의 밀도값은 -1과 1 사이의 값을 가지게 됩니다. 일반적으로, 상관분석에서 0, -1, 1의 값으로 결과가 도출 됩니다.

분석결과 - 1-mode

0	무료수강	발전	온라인	
1	진학	청년	한양대학교	협약
2	개발	공부	교류	기회
3	교육철학	대학원	성장	
4	이념	이상	인터넷	지침기회
5	취득	코로나	협약	협력관계
6	교육	온라인대학	전문가	
7	미래	사이버	연택트	

분석결과 - 밀도값 (1-mode)

	Block_1	Block_2	Block_3	Block_4	Block_5	Block_6	Block_7	Block_8
Block_1	1	1	1	1	1	1	1	1
Block_2	1	1	1	1	1	1	1	1
Block_3	1	1	1	1	1	1	1	1
Block_4	1	1	1	1	1	1	1	1
Block_5	1	1	1	1	1	1	1	1
Block_6	1	1	1	1	1	1	1	1
Block_7	1	1	1	1	1	1	1	1
Block_8	1	1	1	1	1	1	1	1

### ① 결과

각 담론마다 상관관계가 있는 단어들을 군집화 시킴

### ② 블록 밀도값

두 변수의 상관계수의 값이 -1 정반대의 관계를 가짐

두 변수의 상관계수의 값이 0 아무 관계가 없음

두 변수의 상관계수의 값이 1 완벽히 일치한 관계를 가짐

데이터 용량추가 더이엠씨님

문서 내에서 동시에 등장(공출현)하는 단어 사이의 관계를 나타내는 분석으로, 상관관계를 이용하여 단어 간의 관계 패턴에 따라 군집화하는 분석방법입니다.

담론 갯수

2개 4개 8개 16개

분석이 완료되었습니다. 적용

분석단어

- 선택단어(1-mode) 미리보기 다운로드
- 선택 열단어(2-mode) 미리보기 다운로드
- 선택 행단어(2-mode) 미리보기 다운로드

분석결과

- 결과(1-mode) ① 미리보기 다운로드
- 밀도값(1-mode) ② 미리보기 다운로드
- 결과(2-mode) 미리보기 다운로드
- 밀도값(2-mode) 미리보기 다운로드

# 감성분석 하기\_감성사전기반

TEXTOM
감성분석
데이터 용량추기 더아이엠씨님

데이터수집

데이터전처리

수집리스트

정제/형태소분석

**데이터분석**

텍스트마이닝

매트릭스

답론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

72594 / 72594

포털/SNS
뉴스
보유데이터
요청채널
 데이터적용 리스트

	데이터명		생성날짜	용량
<input type="checkbox"/>	테스트		2022-03-03	308 KB
<input type="checkbox"/>	중대재해(카페)		2022-03-03	1011 KB
<input type="checkbox"/>	사스 호텔		2022-03-03	1.87 MB
<input type="checkbox"/>	신종플루 호텔		2022-03-03	3.01 MB
<input type="checkbox"/>	메르스 호텔		2022-03-03	3.42 MB
<input type="checkbox"/>	코로나 호텔		2022-03-03	4.09 MB
<input type="checkbox"/>	호텔감정노동		2022-03-03	285 KB
<input type="checkbox"/>	충북바이오산업		2022-03-03	628 KB
<input type="checkbox"/>	리듬운동 +스포츠		2022-03-03	2.21 MB
<input type="checkbox"/>	메타버스 관광		2022-03-03	2.87 MB

감성분석 사용법 자세히 보기 >>

분석데이터

원문데이터
경제데이터

원문데이터

미리보기
다운로드(Excel)
다운로드(txt)

**분석방법 ①**

감성사전/문서분류 기법을 통해 좀 더 자세한 감성분석을 이용할 수 있습니다.

감성사전
문서분류

**분석결과**

미리보기
다운로드
✓ 적용

미리보기 일부 데이터를 확인 할 수 있습니다. 전체 데이터를 원하는 경우 다운로드 기능을 이용하세요.

감성분석 시각화 결과는 시각화결과 페이지에서 확인하실 수 있습니다.

- ① **분석방법** 원문에서 감성단어의 빈도를 보여주는 '감성사전',  
문서 분류기를 구축하여 긍/부정을 확인할 수 있는 '문서분류' 두 가지 중 분석 방법을 선택합니다.
- ② **사전적용** 적용버튼을 클릭하여 사전기반 감성분석을 진행합니다.

TEXTOM

☰ 감성분석

데이터용량관리

더아이엠씨님

2

원문 데이터

데이터명	생성날짜	용량
미술허 전시장	2022-03-03	991 KB

따뜻한 '빛'의 세계 <빛> 영국 테이트미술관 특별전< 빛>의 인물은 루시마. 지나지는 것을 2005. 테이트미술관 소장 전시장에 들어서면 제일 먼저 백남준(1932-2006)의 유명한 TV 작품이 관람객을 맞는다. 빈 TV 벽스 안에서...

#빛\_영국테이트미술관특별전 #작품소개 #북서울미술관 #빛\_영국테이트미술관특별전 #작품소개 <빛> 영국 테이트미술관 특별전 전시장 도입부는 백남준아트센터에서 소장함...

[전시] 환기미술관 특별기획전 <김환기, 그라다-D...> 독창적 양식의 구축을 향한 김환기의 '그라다-D-슈AW' 환기미술관 2층 전시장 전경 1964년부터 69년의 6년여, 김환기는 작가적 역량을 기반으로 광범위한...

'빛'영국 테이트미술관 특별전@서울시립미술관(SeMA) 9000원 전시장 내 사진 촬영 불가 (이하의 작품사진들은 테이트 미술관의... 북서울미술관 전시장 일구어 서 있던 작품, <그리 레이의 스펙트럼> 유일하게...

국립현대미술관 MMCA 아진회관에서 특별전: 한국미술명작... 사진 전시장 일구어 찾기 위해서, 해명인... 그런데 이렇게 걸으니 너무 좋았어요! 드디어 좋은 미술관 전시장 가는 길 터널을 받을 때 본인확인요...

서울시립 북서울미술관 <빛> 영국 테이트미술관 특별전> 관람 중... <대한민국 여행정보> 카페에서 미술관 후기를 종종 올렸었는데 고고유업으로 이사로 고사전... <도슨트북소리와 전시장에서 다들 다들한 책요 <=> 취사하면 이번 전시는...

게티 이미지 사진전 (예술이전담 환기미술관) 환기미술관 1층 관람시간 오전 10시~오후 7시 매표 및 입장마감 오후 6시 홈페이지 <https:getty4seoul.moodoo.at/getty4seoul 특별안내...>

<사람특별전: Chagall and the Bible> 미디어뮤지엄 사람 특별전: Chagall and the Bible 장르 전시 장소 미디어뮤지엄 기간 2021.11.25. ~ 2022.04.10 상설역 4번... 내 귀에 미술관은 지나뮤지엄 설치하고 로그인해 해아들을 수 있는데요, 조금 번거롭긴 하지만...

[방학특집 1] 프랑스국립자박물관 특별전 강추! (물고기와 배... 프랑스로서 그런지 저어사박물관 미술관 갔어요! 불거리 담~~따 말씀니다. <저는 자연사박물관... 전시장 안에서 다양한 미디어 제물이 있어서 초등 아이까지...>

사람 특별전 이번 사람 특별전에서는 genie 어플을 통해서 '내귀에 미술관' 오디오가이드 콘텐츠를 들으며 작품을... 미디어뮤지엄 사람전의 전시

사람 특별전 Chagall and the Bible 전시 관람 후기... 방문을 철저히 하는 전시장이라 백신훈과 확인 또는 PCR음성검사(48시간이내)를 제출해야만... <직접 전 지켜야하겠죠>^^ 대신 사진은 꼭 찍어 봐주세요!^^

전시/사람특별전 전시 전시장 사진 및 코로나 방역지침에 따라 변경될 수 있습니다. 프라이빗 특별 도슨트 예약은 홈페이지 및 미술관 별도 문의 바랍니다. 오시는길: 서울시 강남구 테헤란로 518 성원센터빌딩 81 B1, 518, Teheran...

전시명물 잘 봐주세요! <사람 특별전: Chagall and 'the Bible'>... 일회 '내귀에 미술관' 미리 들어보고는 괜히 우려를 거둬다 내귀에 미술관의 구약성서 이야기는... 장르 전시 장소 미디어뮤지엄 기간 2021.11.25. ~ ...

사람 특별전 후기... 바로 성우 쌤만님과 함께하는 '내 귀에 미술관' 콘텐츠인데요! 지나 뮤직을 이용하시는 분들이라면... 부모님들까지 전시장을 찾은...

감성분석 결과 (감성사전기반)

구분	빈도(건)	감성강도비율(%)	빈도비율(%)
긍정	2319 / 2822	84.65 / 100.0	82.18 / 100.0
부정	503 / 2822	15.35 / 100.0	17.82 / 100.0

강정 키워드 부정 키워드 세부감성 감성 키워드 클릭 시 해당 키워드가 포함된 원문을 확인할 수 있습니다.

흥미 호감 기쁨

감성분류	빈도(건)	감성강도	빈도 * 감성강도	빈도비율(%)
분뻘하다	190	3.77778	717.7782	6.73
새롭다	41	2.77778	113.8898	1.45
색다르다	30	4.6667	140.001	1.06
깨미있다	19	2.6667	50.6673	0.67
기대하다	17	4.6667	79.33339	0.6
흥미있다	17	3.4444	58.5548	0.6
인상적이다	16	3.4444	55.1104	0.57
신비롭다	14	1.7778	24.8892	0.5
일드적이다	14	3.3333	46.6662	0.5
완하다	12	5.0	60	0.43
이색적이다	10	3.6667	36.667	0.35
일제적이다	8	6.0	48	0.28
	-	-	-	-

1

감성분석 사용법 자세히 보기 >>

분석데이터

원문데이터    정제데이터

원문데이터

미리보기    다운로드(Excel)    다운로드(txt)

분석 방법

감성사전/문서분류 기법을 통해 좀 더 자세한 감성분석을 이용할 수 있습니다.

감성사전    문서분류

분석결과    학습데이터가 적용 되어있습니다.

미리보기    다운로드(txt)    적용

미리보기 일부 데이터를 확인 할 수 있습니다. 전체 데이터를 원하는 경우 다운로드 기능을 이용하세요. 감성분석 시각화 결과는 시각화결과 페이지에서 확인하실 수 있습니다.

## 분석결과 적용되면 분석결과를 확인할 수 있습니다

- 1 감성분석 결과 사전에 기반으로 감성단어를 추출하고 빈도결과를 확인할 수 있습니다.
- 2 원문확인 해당 감성단어가 포함된 원문을 보여줍니다.

※ 감성단어들은 감성분석에 탑재된 형태소 분석기로 출력된 키워드이므로, 텍스트마이닝 페이지에서 직접 편집한 데이터와 차이가 있습니다  
 ※ 감성강도는 세부 감성(흥미, 호감, 기쁨, 통증, 슬픔, 분노, 두려움, 놀람, 거부감) 안에서 표현의 세기를 의미하며, 7점 만점입니다

TEXTOM 감성분석

데이터수집

데이터명검색 1555 / 72610

삭제

포털/SNS 뉴스 보유데이터 요청채널

데이터적용 리스트

감성분석 사용법 자세히 보기 >>

**1** 분석데이터

원문데이터 정제데이터

원문데이터

미리보기 다운로드(Excel) 다운로드(txt)

**분석 방법**

감성사전 문서분류

**2** 학습 데이터

적용

**3** 테스트 데이터

적용

**예시파일 다운로드**

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 업로드해주세요. (비율을 비슷하게 업로드 할 수록 정확한 결과가 나옵니다.)

원문데이터에 직접 극성을 부여한 학습데이터를 업로드하여, 문서의 극성 분석결과를 확인할 수 있습니다.

- 1 분석데이터** 원문데이터 혹은 정제데이터를 선택한 후, 다운로드(Excel)하여 학습데이터를 만들어야 합니다
- 2 학습데이터** 전체 데이터를 분류하기 위한 기준이 되는 데이터로, 각 극성의 비율을 비슷하게 태깅하여 업로드해줍니다
- 3 테스트 데이터** 학습데이터를 기반으로 만들어진 분류기의 성능을 확인하기 위한 테스트 데이터를 업로드해줍니다.

[학습데이터 만드는 법 자세히 알아보기](#)



# 감성분석 하기\_문서분류 기반

TEXTOM
감성분석
데이터용량추가 더이엠씨님

데이터수집

데이터전처리

수집리스트

정제/형태소분석

**데이터분석**

텍스트마이닝

매트릭스

답론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

커스터마이징

감성분석 결과 (문서기반) - 전체

구분	빈도(건)	비율(%)
전체	910	100.0
긍정	527	57.91
중립	351	38.57
부정	32	3.52

분석문장		결과
화이자 코로나 백신 3차 접종 후기! 오늘 코로나 백신 3차 맞고 왔습니다. 주변에서 몸도 약한 사람이 많 그리 빨리... 해서 코로나 1차..	긍정	
위드코로나 사실상 중단 사회적 거리두기 조정 방안 코로나 확진자수가 무섭게 증가하고 있으며 위중증 환자수 역시 매일 늘어가고... 사실상 위..	긍정	
21년 12월 18일(토)~22년 1월 2일(일), 16일간 코로나19... 질병관리청의 코로나19 긴급 위험도 평가 결과, 위험도는 전 국..	긍정	
코로나 자가격리지원금의 신청방법과 금액은 어떻게 될까? 코로나 자가격리 대상일 시 지원금에 대한 이야기를 해볼려고 한다. 얼마나... 그리고..	중립	
인면윤곽 부작용, 코로나 확진 후 발생 주의!!! 인면윤곽 부작용, 코로나 확진 후 발생 주의!!! 오늘은 참 안타까운 이야기 입니다. (먼저..	중립	
코로나19 백신, 3차접종 간격이 3개월로 단축됩니다! 코로나19 백신, 이제 3차까지 맞으셔야 안심입니다! 최근 국내 방 역상황이... 방법..	긍정	
코로나 시국, 마스크 재순환이 가능하다고? 요즘 다시 시작된 코로나 확산으로 걱정이 많으시죠? 코로나19의 장기화로 우 리 사회는 여태껏 겪지..	부정	
코로나 방역 대표업체 확실했어요 가장 먼저 코로나 방역 인증업체인 인허가업체인지부터 살펴보세요 해요. 코로나가 터 진 이후로 전문 방역업체에 ..	중립	

10개

데이터용량추가

적용

예시파일 다운로드

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 업로드해주세요. (비율을 비슷하게 업로드 할 수록 정확한 결과가 나옵니다.)

**1 모델 성능**

정확도	전체	99.34 %
	긍정	97.87 %
	중립	100.0 %
	부정	100.0 %

**분석결과**

전체 (긍정/중립/부정)

**2** 미리보기    다운로드(Excel)    다운로드(txt)

긍정

미리보기    다운로드    **3** 추가분석

중립

미리보기    다운로드    추가분석

부정

- 1 모델 성능** 학습된 모델의 성능을 볼 수 있습니다. 성능이 낮다면 학습데이터를 보정하여 향상시킬 수 있습니다.
- 2 분석결과 미리보기** 학습된 모델을 기반으로 긍/부/중립을 분류하여 결과를 제공합니다.
- 3 추가분석** 분류된 극성(긍정,부정,중립)을 추가분석 구체적인 분석을 진행할 수 있습니다.

# 감성분류기 성능 계산 방법

		예측 결과			
		긍정	중립	부정	정확도
실제 결과	긍정	정답	오답	오답	$\frac{\text{정답}}{\text{정답} + \text{오답} + \text{오답}}$ ※ 긍정 분류기 정확도 계산 방법
	중립	오답	정답	오답	
	부정	오답	오답	정답	

**실제결과** 학습데이터를 이용해 분류된 결과를 의미합니다.(분류모델은 학습데이터를 기반으로 만듭니다.)

**예측결과** 제작한 분류기를 통해 분류된 테스트 데이터의 결과를 의미합니다.

학습데이터를 업로드 한 후 테스트데이터를 별도의 엑셀파일로 만들어 업로드 하면 모델의 성능을 확인할 수 있습니다.

※ 테스트데이터는 학습데이터와 다른 문장으로 구성되어야 정확하게 성능을 측정할 수 있습니다.

성능이 낮다면, 학습데이터를 보완하여 업로드하면 성능을 개선할 수 있습니다.

**분석결과 (텍스트마이닝)**

데이터명	생성날짜	용량
코로나 +메르스	2020-05-12	6.26 MB

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다. [다운로드](#)

그룹	단어	Topic Modeling
1	코로나	277
1	수	73
1	등	71
1	발생	37
1	2015년	27
1	당시	23
1	관련	23
1	전염병	23
1	사람	23
1	3월	20
1	오늘	15
1	위험	9

**Word-level Semantic Clustering**  
문서 내 단어를 임베딩하여 단어 벡터 간의 묶음 관계를 Agglomerative Hierarchical Clustering을 통해 확인하는 방법입니다. 군집 수가 사용자가 지정한 것 보다 작을 경우 임의 지정 됩니다.

1 ▶ 군집 수 (K 값)  개

2 ▶ 군집별 단어 수  개

3 ▶ [적용](#)

**분석결과**

4 ▶ Word-level Semantic Clustering

[미리보기](#) [다운로드\(Excel\)](#) [다운로드\(txt\)](#)

▶ LDA Topic Modeling

[미리보기](#) [다운로드\(Excel\)](#) [다운로드\(txt\)](#)

**Word-level Semantic Clustering** 문서 내 단어들의 공출현 관계를 기준으로 벡터화하여 인접 단어를 같은 군집으로 묶어줍니다

- ❶ **군집 수(K값)** 만들고 싶은 군집의 개수를 적어줍니다
- ❷ **군집별 단어 수** 군집 내에 들어갈 단어의 개수를 적어줍니다
- ❸ **적용** 군집 수(K값)과 군집별 단어 수를 설정 후 적용 버튼을 클릭합니다
- ❹ **분석결과** 적용이 완료되면 연관성이 높은 단어들끼리 군집으로 분류가 된 결과를 확인할 수 있습니다

- 담론분석
- 감성분석
- 토픽분석**
- 시계열분석 (Beta)
- 시각화
- 시각화결과
- 커스터마이징

**분석결과 (텍스트마이닝)**

데이터명	생성날짜	용량
코로나 +메르스	2020-05-12	6.26 MB

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

그룹	단어	Topic Modeling
1	메르스	0.024
1	코로나	0.023
1	사스	0.022
1	것	0.019
1	바이러스	0.015
1	년	0.011
1	19	0.009
1	신종	0.008
1	2015	0.007

**LDA Topic Modeling**  
토픽모델링은 대량의 문서군으로부터 주제(토픽)를 자동으로 찾아내기 위한 알고리즘으로, 유사한 의미를 가진 단어들을 집단화하는 방식으로 토픽을 추천하는 방법입니다.

1 토픽 수  개

2 토픽 단어 수  개

3 랜덤 값  사용  사용안함

샘플링 과정에 포함된 무작위 토픽 할당 기능 사용 여부를 선택합니다. 사용 시 토픽 모델링 결과의 재현성이 떨어질 수 있습니다.

학습데이터가 적용 되어있습니다. 4 적용

**분석결과**

▶ Word-level Semantic Clustering

▶ LDA Topic Modeling

5

**LDA Topic Modeling** 대량의 문서군으로부터 주제(토픽)를 자동으로 찾아내기 위한 알고리즘으로, 유사한 의미를 가진 단어들을 집단화 합니다

- ❶ 토픽 수 만들고 싶은 그룹의 개수를 적어줍니다
- ❷ 토픽 단어 수 그룹 내에 들어갈 단어의 개수를 적어줍니다
- ❸ 랜덤 값 무작위 할당을 진행할 경우 분석 결과의 재현성이 떨어지기 때문에, 같은 데이터로 분석을 진행하더라도 결과값이 달라질 수 있습니다. 분석 결과의 재현성을 확보하고 싶은 경우 **사용안함**을 선택해주세요
- ❹ 적용 토픽 수와 토픽 단어 수, 랜덤 값 설정 후 적용 버튼을 클릭합니다
- ❺ **분석결과** 적용이 완료되면 키워드가 어떤 그룹으로 분류가 되었는지 확인할 수 있습니다

※ Topic Modeling 숫자값은 해당하는 토픽에 단어가 들어갈 확률을 백분율 형태로 표시한 값입니다

담론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

커스터마이징

분석결과 (텍스트마이닝)

데이터명	생성날짜	용량
코로나+메르스	2020-09-22	6.26 MB

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

그룹	확률	단어
7	0.93076235	코로나 검사 후기 코로나검사 위험도와 예방접종으로 코로나 사스 예방을 잘 낫달 내기코로나검사
1	0.94998944	특집 기사 코로나 19 현황 전망 중용 사망 수 자료 따르면코로나19는메르스도 사스 비교 일상 지사 것 후만
1	0.94704944	세계 경제법 워싱턴포스트 전례로 유럽 도미언트로 도입 메르스 재 이어코로나바이러스 쇼 기문사 결과 시간 소모
3	0.9526182	바이러스 전설 이재갑 교수 바이러스 영우 중환자 치료 디와 지사용 메르스 알리코로 우한 우한 계단 조사 결과 사망률
6	0.95262617	재발 저가격의 달 육아 일상 활동 15년 이번코로나19 전세계 국경 차단 상용 물 유행위기 메르스 돌 일 이번
6	0.88749063	코로나 이후 시스템리스산용물류 그래프 같이코로나 이후 전염성
8	0.9571347	공저 중식 회향 황태이 코로나 19 첫 날 검사 중식 회향 황태이 코로나 19 예방 사스 나 메르스 유행 전염성
7	0.94999725	위생마스크 꼭고코로나차지 생마스크 것 위생마스크 꼭고코로나19 예방마스크 착용 마늘 한국인 대표 영양 알고 전세계가메르스 사스와 한국인 마늘
10	0.9608663	2020년 발효종신 대 1학 중간과제를 정리 발송종신 2020년 1학 보건의사소통 중간시험과제물 C 할 코로나19로메르스확산 비교 코로나 감염 원인 수면사립 위험인 식 과가 중용
2	0.95713186	시대 마음 잘 정도 활동 온 코로나 19 방역 대처 수준급 결과 잘만 수 방 메르스사서 미 예방 재난 일
3	0.9624962	산용 코로나 바이러스 개인 구강 위생 예방 과거메르스 사스 바이러스도코로나바이러스 일용 이번코로나19 상용성 나타났습니 메르스 사스 바이러스 유행 데 호비전으로도 상용

분석결과 (텍스트마이닝)

데이터명	생성날짜	용량
코로나+메르스	2020-05-12	6.26 MB

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

그룹	단어	Topic Modeling
1	메르스	0.024
1	코로나	0.023
1	사스	0.022
1	것	0.019
1	바이러스	0.015
1	년	0.011
1	19	0.009
1	산용	0.008
1	2015	0.007

**LDA Topic Modeling**  
토픽모델링은 대량의 문서군으로부터 주제(토픽)를 자동으로 찾아내기 위한 알고리즘으로, 유사한 의미를 가진 단어들을 집단화하는 방식으로 토픽을 추론하는 방법입니다.

- 토픽 수: 10 개
- 토픽 단어 수: 20 개
- 랜덤 값: 사용 / 사용안함
- 적용
- 미리보기
- 미리보기

샘플링 과정에 포함된 무작위 토픽 할당 기능 사용 여부를 선택합니다. 사용 시 토픽 모델링 결과의 재현성이 떨어질 수 있습니다.

학습데이터가 적용 되어있음

**분석결과**

- Word-level Semantic Clustering
  - 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)
- LDA Topic Modeling
  - 5 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)
- LDA Topic Origin Text
  - 5 미리보기
  - 다운로드(Excel)
  - 다운로드(txt)

**LDA Topic Modeling** 대량의 문서군으로부터 주제(토픽)를 자동으로 찾아내기 위한 알고리즘으로, 유사한 의미를 가진 단어들을 집단화 합니다

- 토픽 수** 만들고 싶은 그룹의 개수를 적어줍니다
- 토픽 단어 수** 그룹 내에 들어갈 단어의 개수를 적어줍니다
- 랜덤 값** 무작위 할당을 진행할 경우 분석 결과의 재현성이 떨어지기 때문에, 같은 데이터로 분석을 진행하더라도 결과값이 달라질 수 있습니다. 분석 결과의 재현성을 확보하고 싶은 경우 **사용안함**을 선택해주세요
- 적용** 토픽 수와 토픽 단어 수, 랜덤 값 설정 후 적용 버튼을 클릭합니다
- 분석결과** 적용이 완료되면 키워드가 어떤 그룹으로 분류가 되었는지 확인할 수 있습니다

※ Topic Modeling 숫자값은 해당하는 토픽에 단어가 들어갈 확률을 백분율 형태로 표시한 값입니다

# 시계열분석 하기 (Beta)

TEXTOM
☰
데이터 용량추가 더이엠씨님

수집하기

데이터전처리

수집리스트

정제/형태소분석

**데이터분석**

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

시각화

시각화결과

커스터마이징

요청채널추가

분석단어선택 (시계열분석)

데이터명	생성날짜	용량
한양사이버대학교	2021-03-03	160 KB

단어빈도 TF-IDF

선택단어수  확인 선택단어누락비율  % 다운로드 적용

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

<input type="checkbox"/>	단어	빈도	백분율 (%)	누락비율 (%)
<input type="checkbox"/>	교수	299	1.748%	1.748%
<input type="checkbox"/>	스타벅스	243	1.42%	3.168%
<input type="checkbox"/>	년	217	1.268%	4.436%
<input type="checkbox"/>	대	200	1.169%	5.605%
<input type="checkbox"/>	직원	182	1.064%	6.669%
<input type="checkbox"/>	한양사이버대학교	181	1.058%	7.727%
<input type="checkbox"/>	교육	133	0.777%	8.504%
<input type="checkbox"/>	대학원	119	0.696%	9.2%
<input type="checkbox"/>	한양사이버	107	0.625%	9.825%
<input type="checkbox"/>	모집	100	0.584%	10.41%
<input type="checkbox"/>	국내	98	0.573%	10.983%
<input type="checkbox"/>	등	90	0.526%	11.509%
<input type="checkbox"/>	중	81	0.473%	11.982%

수집량 시각화

분석단어	수집날짜	용량
한양사이버대학교	2021-02-19	160 KB

다운로드 초기화

전체 내이비 다음 구글

소집단위에서 설정한 기간별로, 선택한 단어의 출현 빈도를 분석할 수 있습니다. 시계열분석은 수집하기에서 소집단위를 사용하여 수집한 데이터만 분석이 가능합니다.

분석단어 선택

1 바로선택하기

시계열분석 결과

기간별 수집량

2 기간별 수집량 시각화

기간별 단어빈도

미리보기 다운로드(Excel) 다운로드(txt)

단어빈도 시각화

단어별 수집량 시각화

- ❶ 분석단어 선택 분석하고 싶은 상위 단어 10개정도 적용시켜줍니다.
- ❷ 기간별 수집량 기간별로 수집한 각 채널의 수집량을 시각화 차트로 보여줍니다.

※ 수집단위를 이용하여 수집을 해야지만 시계열 분석 사용이 가능합니다.

# 시계열분석 하기 (Beta)

수집하기

- 데이터전처리
- 수집리스트
- 정제/형태소분석
- 데이터분석**
- 텍스트마이닝
- 매트릭스
- 담론분석
- 감성분석
- 토픽분석
- 시계열분석 (Beta)
- 시각화
- 시각화결과
- 커스터마이징

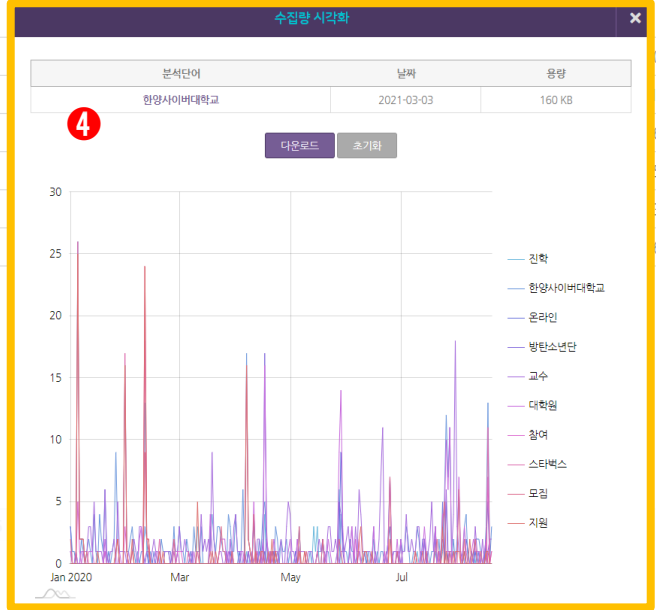
**3**

분석결과 (시계열분석)

데이터명	생성날짜	용량
한양사이버대학교	2021-03-03	160 KB

단위수집된 데이터만 확인 할 수 있습니다. 다운로드

단어	날짜	빈도
친학	2020-10-13	2
친학	2020-05-18	3
친학	2020-01-13	2
친학	2020-05-27	3
친학	2020-08-19	3
친학	2020-11-25	1
친학	2020-01-10	2
친학	2020-03-18	13
친학	2020-09-28	1
친학	2020-05-19	1
친학	2020-08-18	1
친학	2020-12-27	1
친학	2020-03-19	1
친학	2020-09-01	1
친학	2020-07-16	1



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

분석단어 선택

**시계열분석 결과**

▶ 기간별 수집량

▶ 기간별 단어빈도

**3** 미리보기       

▶ 단어빈도 시각화

**4**

- 3** 기간별 단어빈도 기간별 단어의 빈도를 보여줍니다.
- 4** 단어빈도 시각화 기간별 단어빈도를 시각화 차트로 보여줍니다.

※ 수집단위를 이용하여 수집을 해야지만 시계열 분석 사용이 가능합니다.

# 시각화 결과 확인하기

The screenshot shows the TEXTOM interface with a word cloud visualization and a settings panel on the right. The word cloud features the word '바이러스' (virus) in large purple letters, surrounded by other terms like '사스' (SARS), '백신' (vaccine), '신종플루' (new influenza), and '확산' (spread). The settings panel on the right includes a '다운로드' (download) button, a '클라우드 모양 변경' (change cloud shape) section with a '샘플이미지1' and '샘플이미지2' input, a '색상선택' (color selection) section with a color palette, and a '키워드 선정' (keyword selection) section with a table of keywords and their frequencies.

빈도	키워드	내림차순	오름차순	
상위 30개	선택해제	숫자	영어	
1~50	51~100			
<input type="checkbox"/>	메르스	296	<input checked="" type="checkbox"/> 증상	23
<input type="checkbox"/>	코로나	277	<input type="checkbox"/> 현재	23
<input checked="" type="checkbox"/>	사스	248	<input checked="" type="checkbox"/> 후보물질	23
<input checked="" type="checkbox"/>	바이러스	238	<input type="checkbox"/> 말	22
<input type="checkbox"/>	것	129	<input checked="" type="checkbox"/> 슈펙트	22
<input type="checkbox"/>	년	110	<input type="checkbox"/> 일	22

워드클라우드에는 문서의 키워드를 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 시각화로, 단어빈도 결과값을 사용합니다

① **색상선택** 4가지의 기본 색상 조합 또는 우측의 상, 중, 하에서 직접 원하는 색상을 선택할 수 있습니다

② **키워드 선정** 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다

※ 상위 빈도 순으로 최대한 유의미한 키워드들을 선택하는 것이 좋습니다

※ 선택한 키워드 중 숫자나 영어가 있을 시, 숫자 또는 영어 버튼을 클릭하면 해당 키워드가 강조되며 다른 키워드들은 회색조 처리가 됩니다

③ **클라우드 모양 변경** 배경이 없는 jpg, png 이미지를 업로드하면 업로드한 모양의 워드클라우드가 만들어집니다



# 시각화 결과 확인하기

TEXTOM
시각화결과
데이터용량추가 더아이멤버십

수집하기
워드클라우드
바차트
에고네트워크
네트워크
개체명인식
LDA
클러스터링
매트릭스 차트
담론분석 (1-mode)
담론분석 (2-mode)
문서 감성분석
단어 감성분석
감성단어 분석
감성단어 워드클라우드

데이터전처리

수집리스트

정제/형태소분석

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

**시각화**

시각화결과

커스터마이징

분석단어	수집날짜	용량
코로나+메르스	2020-03-20	6.26 KB

3 다운로드

1 시각화 설정  
빈도 백분율 데이터 표시

2 키워드 선정  
빈도 키워드 내림차순 오름차순

상위 25개 선택해제

1~25 26~50

<input type="checkbox"/> 19	103	<input checked="" type="checkbox"/> 사람	23
<input type="checkbox"/> 2015	29	<input checked="" type="checkbox"/> 사스	248
<input type="checkbox"/> 2015년	27	<input checked="" type="checkbox"/> 사태	63
<input type="checkbox"/> 간	35	<input checked="" type="checkbox"/> 세계	55
<input checked="" type="checkbox"/> 감염	52	<input type="checkbox"/> 수	73
<input type="checkbox"/> 감염병	47	<input checked="" type="checkbox"/> 신종	60
<input type="checkbox"/> 감염증	24	<input checked="" type="checkbox"/> 신종플루	65
<input checked="" type="checkbox"/> 개발	81	<input checked="" type="checkbox"/> 유행	25
<input type="checkbox"/> 것	129	<input type="checkbox"/> 이번	24
<input type="checkbox"/> 경우	24	<input type="checkbox"/> 이후	25
<input type="checkbox"/> 과거	32	<input checked="" type="checkbox"/> 일일악플	39
<input type="checkbox"/> 관려	22	<input type="checkbox"/> 처	42

바차트는 문서의 단어빈도를 빈도에 비례하는 길이의 막대로 나타낸 차트입니다

① **시각화 설정** 빈도는 막대의 색상을, 백분율은 점의 색상을 선택할 수 있으며, 데이터 표시는 차트 위 빈도, 백분율 값 표시 여부를 선택할 수 있습니다

② **키워드 선정** 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다

※ 상위 빈도 순으로 최대한 유의미한 키워드들을 선택하는 것이 좋습니다

※ 막대 위에 마우스를 올리시면 데이터 값을 확인할 수 있습니다

# 시각화 결과 확인하기

TEXTOM
시각화결과
데이터용량추가 더어이씨님

수집하기
워드클라우드 바차트 에고네트워크 네트워크 개체명인식 LDA 클러스터링 매트릭스 차트 담론분석 (1-mode) 담론분석 (2-mode) 문서 감성분석 단어 감성분석 감성단어 분석 감성단어 워드클라우드

분석단어	수집날짜	용량
코로나+메르스	2020-03-20	6.26 KB

다운로드

**1** 색상선택  
배경색 상 중상 중  
글자색 상 중상 중 중하 하

**2** 키워드 선정  
빈도  키워드  내림차순  오름차순

선택한 키워드: 25개

키워드	빈도
메르스	296
코로나	277
사스	248
바이러스	238
것	129
년	110
19	103
때	92
개발	81
수	73
등	71
신종플루	65

선택한 키워드: 25개

키워드	빈도
발생	37
치료	37
간	35
중	35
과거	32
대응	31
효과	29
2015	29
2015년	27
때문	27
국내	26
답변	25

에고네트워크는 문서의 단어빈도를 빈도에 비례하는 크기의 원으로 나타낸 시각화입니다

**1** 색상선택 배경색은 원의 색상을, 글자색은 키워드의 색상을 선택할 수 있습니다

**2** 키워드 선정 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다

※ 상위 빈도 순으로 최대한 유의미한 키워드들을 선택하는 것이 좋습니다

※ 원들이 한곳에 모여져 겹쳐서 보이는 경우, 가운데 원을 잡고 흔들어 주시면 고르게 분산됩니다

# 시각화 결과 확인하기

TEXTOM 시각화결과 데이터용량추기 더아이엠씨님

수집하기 워드클라우드 바차트 에고네트워크 **네트워크** 개체명인식 LDA 클러스터링 매트릭스 차트 답론분석 (1-mode) 답론분석 (2-mode) 문서 감성분석 단어 감성분석 감성단어 분석 감성단어 워드클라우드

분석단어	수집날짜	용량
코로나+메르스	2020-03-20	6.26 KB

**3** 다운로드

**1** 시각화 설정  
 선 색상 ● 화살표 모양 1 2 3 4

**2** 키워드 선정

빈도	키워드	내림차순	오름차순
✓ 상위 50개	선택해제		
<input checked="" type="checkbox"/>	메르스	백신	6
<input checked="" type="checkbox"/>	비상	경제시국	6
<input checked="" type="checkbox"/>	대한민국	방역체계	6
<input type="checkbox"/>	답변서스	메르스	6
<input type="checkbox"/>	사태	당시	6
<input checked="" type="checkbox"/>	감염병	대중	6
<input type="checkbox"/>	말배	해당할정도	6
<input checked="" type="checkbox"/>	바이러스	감염	6
<input type="checkbox"/>	치사	감염	6
<input type="checkbox"/>	19	신종플루	6
<input type="checkbox"/>	증상	것처럼코로나19	6
<input type="checkbox"/>	2020년	3월	6
<input type="checkbox"/>	세계	대유행	6
<input type="checkbox"/>	메르스	중증호흡기중후군	6
<input type="checkbox"/>	답변	바이러스	6

네트워크는 점과 화살표로 단어 간의 연결관계를 나타낸 시각화로, N-gram 결과값을 사용합니다

- ❶ 시각화 설정 선 색상은 화살표의 색상을, 화살표 모양은 화살표의 머리 모양을 선택할 수 있습니다
- ❷ 키워드 선정 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다

※ 상위 빈도 순으로 조사를 제외한 최대한 유의미한 키워드 조합들을 선택하는 것이 좋습니다

※ 원문에서 한 단어였었던 키워드들이 각각의 키워드로 많이 나타난다면, 데이터 편집을 더 해주셔야 합니다

# 시각화 결과 확인하기



개체명인식(트리맵)은 형태소 분석한 데이터를 미리 정의된 14개의 개체명 범주로 분류하여 트리맵으로 나타낸 시각화입니다

① **개체명** 사람, 학문, 대상물, 기관, 지역, 문명, 날짜, 시간 숫자, 사건/사고, 동물, 식물, 금속, 용어 [개체명 인식 자세히 알아보기](#)

※ 개체명 인식기에 기본 탑재된 형태소 분석기로 출력된 키워드이므로, 텍스트마이닝 페이지에서 직접 편집한 데이터와는 차이가 있습니다

※ 해당 문서에 14개 개체명에 속하는 키워드가 없다면 개체명이 나타나지 않습니다



# 시각화 결과 확인하기

TEXTOM
시각화결과
데이터 용량추기 더이엑스님

수집하기
워드클라우드
바차트
에고네트워크
네트워크
개체명인식
LDA
클러스터링
매트릭스 차트
담론분석 (1-mode)
담론분석 (2-mode)
문서 감성분석
단어 감성분석
감성단어 분석
감성단어 워드클라우드

분석단어	수집날짜	용량
한양사이버대학교	2021-02-19	160 KB

다운로드

담론분석의 시각화는 매트릭스 데이터(공출현 빈도)의 결과값을 사용합니다. (담론 분석을 진행해야 시각화 결과를 확인할 수 있습니다)

- ① 군집 설정할 수 있습니다
- ② 시각화 설정 군집을 이동하여 위치를 변경 할 수 있습니다
- ③ 1-mode, 2-mode를 분석하여 각각 시각화 결과를 볼 수 있습니다

# 시각화 결과 확인하기

TEXTOM
시각화결과
데이터 용량추가 더이엠씨님

수집하기

---

데이터전처리

수집리스트

정제/형태소분석

---

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

---

시각화

시각화결과

커스터마이징

워드클라우드   바차트   에고네트워크   네트워크   개체명인식   LDA   **클러스터링**   매트릭스 차트   담론분석 (1-mode)   담론분석 (2-mode)   문서 감성분석   단어 감성분석   감성단어 분석   감성단어 워드클라우드

분석단어	수집날짜	용량
코로나 + 메르스	2020-03-20	6.26 KB

[다운로드](#)

전체도픽

클러스터링은 Word-level Semantic Clustering 분석을 트리맵으로 나타낸 시각화입니다  
(Word-level Semantic Clustering 분석을 진행해야 시각화 결과를 확인할 수 있습니다)

- ※ 전체도픽 위에 마우스를 올리면 해당 토픽에 속해있는 키워드들의 빈도 총합을 확인할 수 있습니다
- ※ 토픽 키워드를 클릭하면 해당 토픽에 속한 키워드를 확인할 수 있습니다
- ※ 토픽분석 페이지에서 개수를 적용하지 않으면 시각화가 나타나지 않습니다

# 시각화 결과 확인하기

TEXTOM

시각화결과

데이터용량추가 더이앤티님

수집하기

워드클라우드   바차트   에고네트워크   네트워크   개체명인식   **LDA**   클러스터링   매트릭스 차트   담론분석 (1-mode)   담론분석 (2-mode)   문서 감성분석   단어 감성분석   감성단어 분석   감성단어 워드클라우드

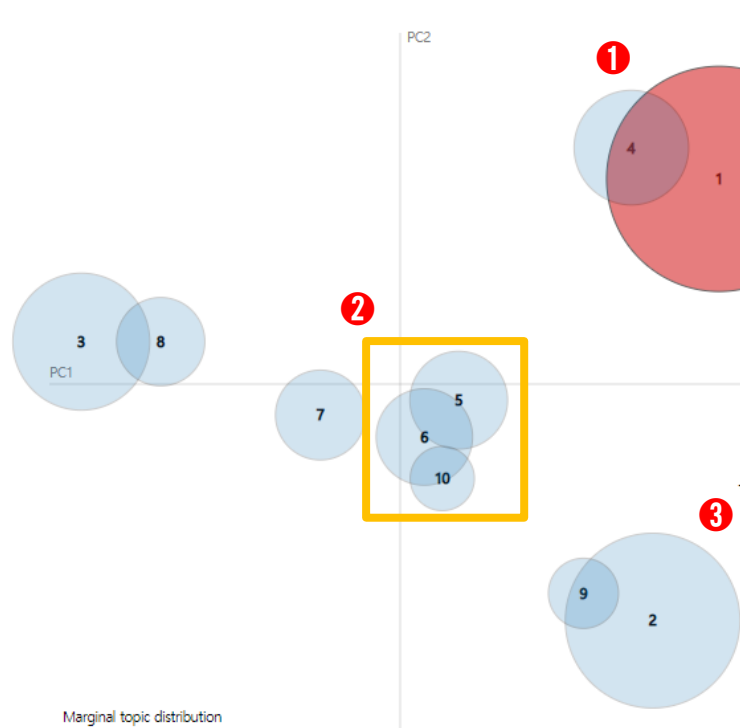
Selected Topics: 1   Previous Topic   Next Topic   Clear Topic

4 Slide to adjust relevance metric<sup>(2)</sup>

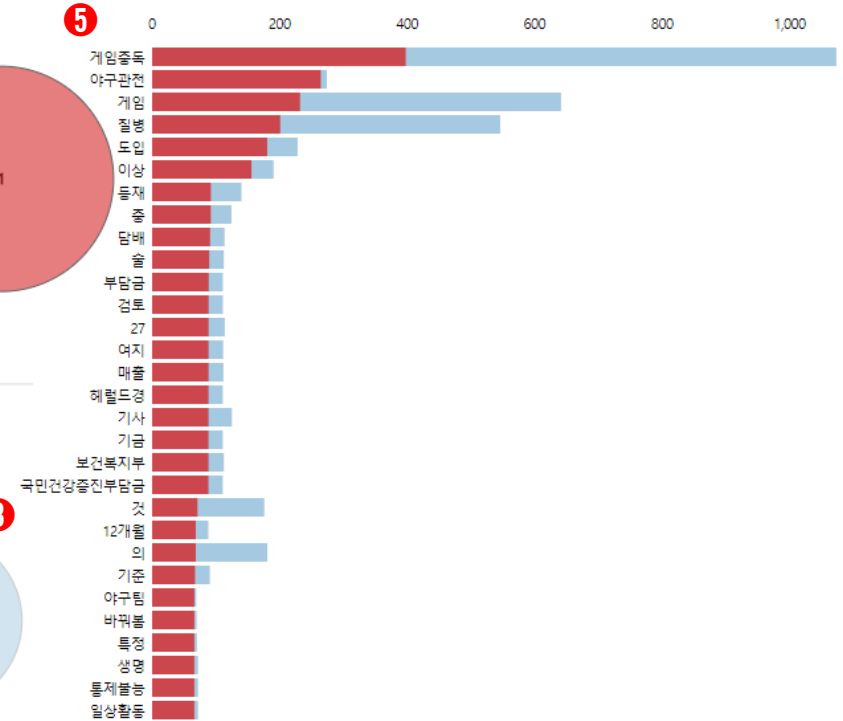
$\lambda = 1$

0.0   0.2   0.4   0.6   0.8   1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (32.3% of tokens)



LDA는 LDA Topic Modeling 분석의 결과값을 시각화로 제공합니다.  
(LDA Topic Modeling 분석을 진행해야 시각화 결과를 확인할 수 있습니다)



## LDA Topic Modeling 시각화 해석

- ① 토픽선택** 토픽분포도에서 토픽을 클릭하거나 토픽번호를 'Selected Topic'에 직접 입력하여 토픽을 선택하면 토픽을 구성하는 30개(사용자 설정 개수) 단어를 확인할 수 있습니다
- ② 토픽 간의 거리** 토픽 간의 거리가 멀 수록 판별 타당도가 높고 주제가 뚜렷하게 구분됩니다  
토픽 간의 거리가 가깝거나 겹쳐져 있다면 판별 타당도가 낮음으로 비슷한 주제를 나타냅니다
- ③ 토픽의 크기** 토픽 원의 크기가 클 수록 높은 빈도수의 단어들로 구성되어 있습니다  
가장 큰 원이 메인 토픽이라고 해석할 수 있습니다
- ④  $\lambda$ (람다) 값 설정**  $\lambda$ (람다) 값을 조절하여 토픽을 구성하는 단어의 출현 조건을 설정할 수 있습니다
- ⑤ 토픽 구성 단어** 토픽을 구성하는 단어들을 확인할 수 있으며, 파란 막대그래프는 전체 단어의 빈도를, 빨간 막대그래프는 해당 토픽에서의 빈도를 보여줍니다

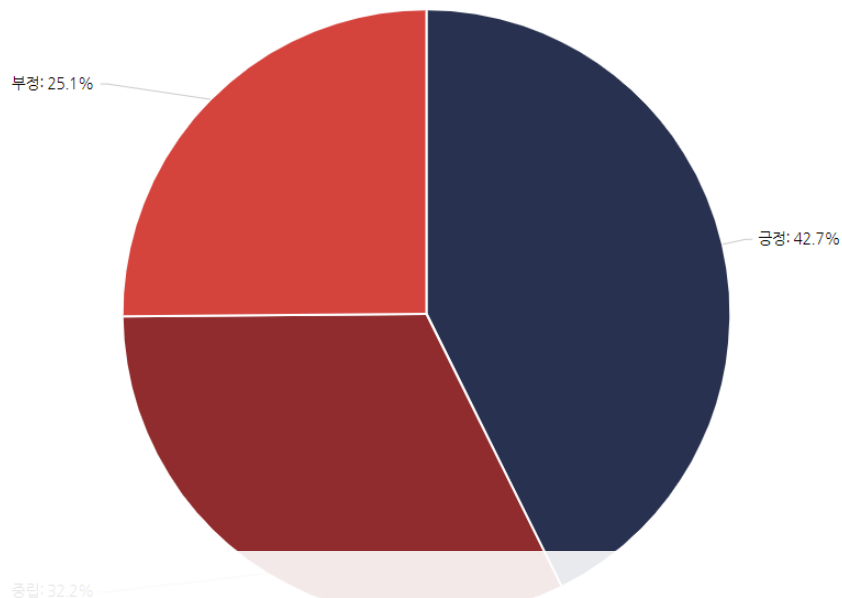
# 시각화 결과 확인하기

수집하기

워드클라우드   바차트   에고네트워크   네트워크   개체명인식   LDA   클러스터링   매트릭스 차트   담론분석 (1-mode)   담론분석 (2-mode)   **문서 감성분석**   단어 감성분석   감성단어 분석   감성단어 워드클라우드

분석단어	수집날짜	용량
트로트 프로그램	2021-02-08	1.12 MB

다운로드



문서 감성분석은 원문데이터 전체를 업로드한 학습데이터 기준으로 극성을 분류하여 파이 차트로 나타낸 시각화입니다

※ 차트 위에 마우스를 올리시면 문서 건수와 백분율 값을 확인할 수 있습니다

※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

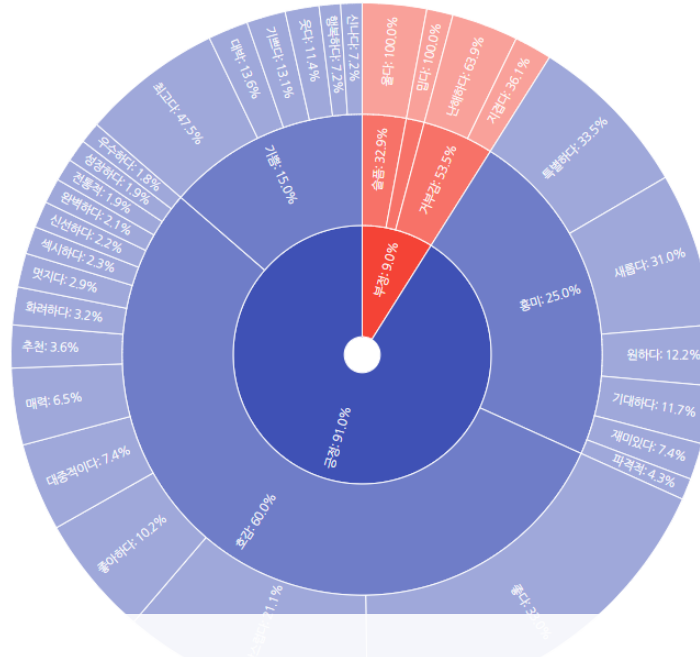
# 시각화 결과 확인하기

워드클라우드 | 바차트 | 에고네트워크 | 네트워크 | 개체명인식 | LDA | 클러스터링 | 매트릭스 차트 | 담론분석 (1-mode) | 담론분석 (2-mode) | 문서 감성분석 | **단어 감성분석** | 감성단어 분석 | 감성단어 워드클라우드

수집하기

분석단어	수집날짜	용량
트로트 프로그램	2021-02-08	1.12 MB

다운로드



키워드 감성분석은 원문데이터에 있는 감성과 관련된 키워드의 감성강도를 파이 차트로 나타낸 시각화입니다

※ 차트 위에 마우스를 올리시면 감성강도의 백분율 값을 확인할 수 있습니다

※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

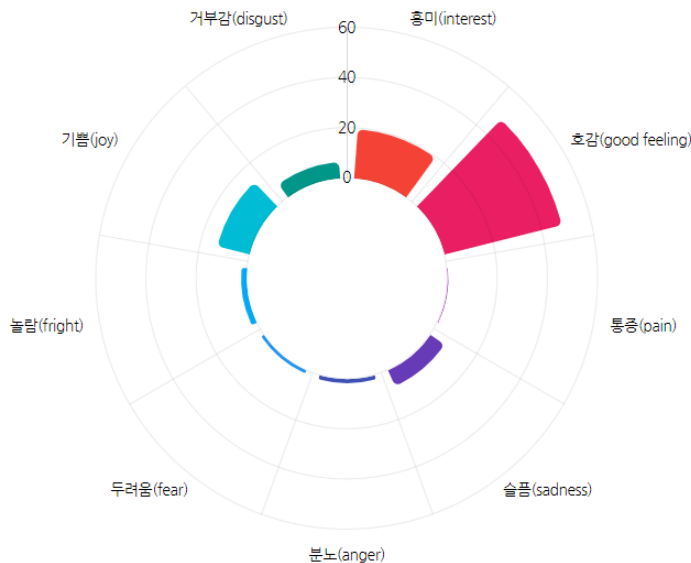
# 시각화 결과 확인하기

수집하기

워드클라우드   바차트   에고네트워크   네트워크   개체명인식   LDA   클러스터링   매트릭스 차트   담론분석 (1-mode)   담론분석 (2-mode)   문서 감성분석   단어 감성분석   감성단어 분석   감성단어 워드클라우드

분석단어	수집날짜	용량
트로트 프로그램	2021-02-08	1.12 MB

다운로드



어휘 감성분석은 원문데이터에 있는 감성과 관련된 키워드의 세부감성비율을 레이다 차트로 나타낸 시각화입니다

※ 차트 계열 위에 마우스를 올리시면 세부감성비율 값을 확인할 수 있습니다

※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

# 시각화 결과 확인하기

TEXTOM
시각화결과
데이터 용량추가 더아엠씨님

수집하기

---

데이터전처리

수집리스트

정제/형태소분석

---

데이터분석

텍스트마이닝

매트릭스

담론분석

감성분석

토픽분석

시계열분석 (Beta)

---

**시각화**

시각화결과

커스터마이징

워드클라우드   바차트   에고네트워크   네트워크   개체명인식   LDA   클러스터링   매트릭스 차트   담론분석 (1-mode)   담론분석 (2-mode)   문서 감성분석   단어 감성분석   감성단어 분석   감성단어 워드클라우드

분석단어	수집날짜	용량
이주1 다문화	2020-09-21	1.32 KB

다운로드

클라우드 모양 변경

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

샘플이미지1, 샘플이미지2

1 색상선택

2 키워드 선정

빈도   키워드   내림차순   오름차순

상위 30개   선택해제   숫자   영어

1~50		51~100		
<input checked="" type="checkbox"/>	울다	136	<input type="checkbox"/> 우아하다	6
<input checked="" type="checkbox"/>	안정적이다	131	<input type="checkbox"/> 풍성하다	6
<input checked="" type="checkbox"/>	사랑스럽다	96	<input type="checkbox"/> 걱정하다	5
<input checked="" type="checkbox"/>	전통적	94	<input type="checkbox"/> 고취되다	5
<input checked="" type="checkbox"/>	어렵다	90	<input type="checkbox"/> 부엌다	5
<input checked="" type="checkbox"/>	행복하다	87	<input type="checkbox"/> 불안	5
<input checked="" type="checkbox"/>	원하다	83	<input type="checkbox"/> 서투르다	5

감성단어 워드클라우드는 감성단어 빈도분석의 감성 키워드의 결과값을 사용합니다.

- ① **색상선택** 4가지의 기본 색상 조합 또는 우측의 상, 중, 하에서 직접 원하는 색상을 선택할 수 있습니다
- ② **키워드 선정** 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다
- ③ **클라우드 모양 변경** 배경이 없는 jpg, png 이미지를 업로드하면 업로드한 모양의 워드클라우드가 만들어집니다

# 시각화 커스터마이징하기

TEXTOM 커스터마이징 데이터용량추가 더아이멤버십

수집하기

데이터전처리  
수집리스트  
정제/형태소분석

데이터분석  
텍스트마이닝  
매트릭스  
담론분석  
감성분석  
토픽분석  
시계열분석 (Beta)

시각화  
시각화결과  
커스터마이징

워드클라우드 바차트 에고네트워크 파이차트 라인차트 N-gram 네트워크 1-way 워드트리 트리맵 2-way 워드트리

워드클라우드

A열 : 키워드 (100개)  
B열 : 빈도

결과보기

코로나+메르스\_빈도100.xlsx

아래 이미지와 같은 형식으로 된 빈도 엑셀 파일을 업로드 하시면 최대 100개 단어를 반영하는 워드 클라우드가 생성됩니다.

Sample						
	A	B	C	D	E	F
1	빅데이터	29871				
2	분석	6118				
3	활용	3456				
4	경보	2647				
5	전문가	2408				
6	기술	2086				
7	산업	1911				
8	기반	1576				
9	연구	1544				
10	교육	1431				

시각화 커스터마이징은 무료이며, 9개의 각 시각화별 양식에 맞는 파일을 업로드만 하면 원하는 시각화 결과물을 바로 확인할 수 있는 기능입니다

① **파일업로드** 좌측 네모 박스 안의 정보와 우측 네모 박스 안의 Sample 데이터를 참고하여 각 시각화에 맞게 적합한 데이터를 엑셀 파일로 업로드해줍니다

② **결과보기** 업로드한 파일의 시각화 결과물을 확인할 수 있습니다