빅데이터 분석 솔루션 TEXTOM 매뉴얼

빅데이터 수집에서 시각화까지!

v4.0

TEXTOM

-

۲

7

 (\mathbf{r})

한눈에 보는 TEXTOM 프로세스

방대하고 복잡한 텍스트 자료,

이제 **효율적으로 분석**하여

연구, 마케팅, 여론분석 등 다양한 곳에 효과적으로 활용해보세요!



Contents



- 1. <u>데이터 수집하기</u>
- 2. <u>데이터 전처리하기</u>
- 3. <u>데이터 분석하기</u>
- 4. <u>데이터 시각화하기</u>

클릭 시 해당 단계 첫 페이지로 이동합니다

회원가입 / 로그인



Start TEXTOM 을 눌러 회원가입&로그인 페이지로 이동합니다

회원가입 증명서류 및 무료데이터용량 지급 관련 안내 바로가기

가입 승인은 영업일 기준 1~3일 이내에 완료됩니다

가입 승인이 완료되면 로그인 후 솔루션 사용이 바로 가능합니다

TEXTOM 텍스톰을 이용하시려면 로그인이 필요합니다.	TEXTOM 더아이엠씨 님의 접속을 환영합니다!
아이디를 있려해주세요	START
비밀번호를 입력하세요	TEXTOM 관리 \longrightarrow 게시판 관리 \longrightarrow MY PAGE \longrightarrow LOGOUT \longrightarrow
로그인 아이디/비밀번호찾기→ <u>회원가입하기</u> →	

회원가입 / 로그인

로그인 시 첫 화면

글 수집하기

ΤΕΧΤΟΜ

<u>데이터 용량추가</u> 더아이엠씨님 🗸 😍

■ 수집하기	포털/SNS	뉴스 보유데이터 요청채널	
▲ 데이터전처리	키워드	키워드추가를 사용하면 여러개의 <u>수집리스트를</u> 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성) 전체수집 요약수집	৩ ই/]জঁ
Ⅲ 데이터분석 × 텍스트마이닝 매트릭스 감성분석	기간	 값 2020-03-09 ~ 🗰 2020-03-16 내이버 학술정보전체, 다음 웹문서, 트위터, 페이스복, 유튜브는 기간 설정이 불가능합니다. 	ছি/ছে ৫
토픽분석 교 시각화 ·	수집단위	사용 사용안함 채널별로 최대 1,000건의 문서를 수집합니다.	
시각화결과 커스터마이징	채널 수집정보	NAVER 네이버 전체 블로그 카페 지식IN 뉴스 웹문서 로 학술경보전체 다음 전체 블로그 카페 뉴스 웹문서 도 도	
	_	Google 구글 전체 뉴스 구글페이스북 웹문서	
4.03/30/14271		수집하기 페이지에서는 분석할 데이터를 수집 하거나 가지고 있는 데이터를 업로드 할 수 있습니다] 페이스북
+ 요정채널주가		이 했네 이 다른 했더이 모셔르 스지했고가 했는 겨오에는 '이처했더'은 토했 이리 한 스 이스니다. 이처했더 바르기기 🔪	-



техтом	로 수집하기 	<u>데이터 용량추가</u> 더아이엠씨님 🗸 💄
🛢 수집하기	1 2 3 4 포털/SNS 뉴스 보유데이터 요청채널	
▲ 데이터전처리 ×	키워드 키워드추가를 사용하면 여러개의 <u>수집리스트를</u> 한번에 생성할 수 있습니다. (동일한 수집조건, 다른 키워드의 리스트로 생성)	হ হিন্দু হ
정제/형태소분석	전체수집 요약수집 ?	
≔ 분석리스트 🗸		

● 포털/SNS <u>네이버, 다음, 구글, 바이두, 유튜브, 트위터, 페이스북</u> 데이터 수집이 가능합니다

❷ 뉴스 <u>KBS, MBC, SBS, YTN, 조선일보, 중앙일보, 동아일보, 한겨레, 경향신문 등</u> 언론사 20곳의 기사 데이터 수집이 가능합니다.

❸ 보유데이터 <u>텍스트로 작성된 pdf, txt, xlsx</u> 형식의 파일 업로드가 가능합니다

④ 요청채널 텍스톰이 제공하는 채널 외 특정 사이트의 데이터 수집이 필요할 때, 별도의 추가 비용을 지불하고 이용 가능합니다.

※ 데이터 수집은 데이터 용량이 차감되지 않습니다

1. 수집하기

수집 키워드

ΤΕΧΤΟΜ	E	수집하기	데이터 용량추가 더아이엠씨님 🗸 😍
을 수집하기	포털/	SNS 뉴스 보유데이터 요청채널	
▲ 데이터전처리 · ·	1 गभ	2 키워드추가 연산자 연산자 연산자	হিিহে ৫
정제/형태소분석		4 전체수집 요약수집 ②	

● 키워드 란에 기입한 형태 그대로, 각 채널 검색창에 검색을 하고 그 검색 결과를 수집하게 됩니다.

- 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
- 키워드 추가 기능은 여러 키워드를 동일한 설정(기간, 수집단위, 채널)으로 수집하고 싶을 때 사용하면
- 유용한 기능입니다
- ④ 전체수집, 요약수집 설정은 데이터의 본문 영역을 어떻게 수집할 지를 정하는 기능입니다.

전체수집, 요약수집 추가 설명 바로가기

※ 전체수집 시 키워드와 관련 없는 스팸 데이터가 많습니다. 왜 그런가요? 각 사이트의 페이지마다 레이아웃이 모두 일관적이지 않아 본문 근처의 리스트, 광고 등의 텍스트가 모두 수집되는 경우가 많습니다. 본문 전체 내용이 꼭 필요한 경우가 아니라면, 요약수집을 사용해 주셔도 키워드에 대한 분석은 충분히 가능합니다.

1. 수집하기

수집 기간 / 단위

≔ 데이터분석	~							
텍스트마이닝		1 기간	2020-04-14	~ 🗰 2020-04-21	1주	3개월 1년	1	2 초기화
매트릭스			네이버 학술정보전체, 다음 웹문서, 트위	터, 페이스북, 유튜브는 기간 설정이 불가능합니다.				
감성분석								
토픽분석		2 수집단위	사용 사용안함					
區 시각화	~		● 일 ○ 주 ○ 월 (○ 년				
시각화결과			채널별로 최대 1,000건의 문서를 수집합 예) 2017.09.01 ~ 2018.02.28 기간의	합니다. 문서를				
커스터마이징			'일' 단위 수집(181일): 채널(섹션)별로 '주' 단위 수집(27주)- 채널(섹션)별로 초 '월' 단위 수집(7달)- 채널(섹션)별로 최대 '년' 단위 수집(2년)- 채널(섹션)별로 최대	최대 181,000건의 문서 수집, 181개의 수집리스트 생성 리대 27,000 건의 문서 수집, 27개의 수집리스트 생성 대 7,000건의 문서 수집, 7개의 수집리스트 생성 대 2,000건의 문서 수집, 2개의 수집리스트 생성				

● 설정한 기간 내에 만들어진 데이터를 수집하게 됩니다

※ 포털/SNS의 **네이버 학술정보전체, 다음 웹문서, 유튜브, 트위터, 페이스북**은 <u>기간 설정이 적용되지 않으며,</u>

<u>수집이 진행되는 시점의 해당 채널 검색 결과를 그대로 수집하게 됩니다</u>

※ 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 <u>네이버 블로그, 카페, 지식IN, 뉴스 / 다음 블로그, 카페, 뉴스 / 구글 뉴스</u>입니다

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

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

♪ 포털/SNS	채널 수집정보	NAVER	네이버 전체 블로그	▶카페 지식IN 뉴스 웹문서	학술정보전체	
		Ddłm	다음 전체 블로그	□ 카페 □ 뉴스 □ 웹문서		
		Google	- 구글 전체 - 뉴스	☐ 구글페이스북		
		Bai <mark>然</mark> 百度	바이두	You Tube	twitter 🗌 म्रास	facebook 미미스북
2 뉴스	채널	뉴스 채널은 수집단위	를 사용할 수 없습니다. (채널벌수집정보 : 제·	목, 본문, URL, 날짜)		
		KBS 😂	KBS	MBC	SBS SBS	•• YTN
		CosantMeda 조선일보	조선일보	· 중앙일보 중앙일보	중아일보 등아일보	한[21 21
		경향신문	경향신문	한국일보 Tiel Kella Tielt	'서울신문 서울신문	⑦연압뉴스
			NEWS1	NEWSIS	Oh. my. News!	노켓뉴스
		매일경제	매일경제	한국경제 한 국경제	*전자신문 전자신문	ZDNet Korea



원하는 데이터만 지정할 수 있는 유용한 기능입니다

②, ③ 시트지정 및 컬럼지정 기능은 엑셀 파일 업로드 시, 많은 시트와 열들 중에서

※ pdf 파일 내 이미지는 인식하지 않습니다

※ txt 파일은 UTF-8 인코딩 파일로 업로드 해 주셔야 합니다 UTF-8 인코딩 설정방법 참고 ▶

※ 파일 형식은 txt, pdf, x

가지고 있는 데이터를

txt : UTF-8 인코딩 파일로 업로드합니다.		
✓ 첫 번째 두 번째 세 번째 직접지경: 적용 할 <u>시트를</u> 지정합니다. 직접지경 : 예) 4,5,6		
✓ A열 B열 C열 D열 E열 F열 직접지정: 적용 할 <u>권령을</u> 지정합니다. 직접지정 : 예) D,G,H		
	다른 이름으로 처장	×
네이터를 텍스톰에 업로드하여 정제, 분석, 시각화 할 수 있습니다 xt, pdf, xlsx, xls 파일만 가능합니다	 ← → <	 ► ► ► ► ► ■

15' USB 드라이브 (F:) 🥔 네트워크

> 파일 이름(N): *.txt 파일 형식(T): 텍스트 문서(*.txt)

< 플더 승기7

				_	_
техтом	E	수집하기	<u>데이터 용량추가</u> 더아	이엠씨님 🗸	8
■ 수집하기	포털/9	NS 뉴스 보유데이터 요청채널 [®]			
▲ 데이터전처리 🛛 🗸	_				
	1 데이터	업로드			=
수십리스트	-	txt,.pdf,xls,xlsx 파일만 업로드 가능합니다.			
정제/형태소분석					
∷ 데이터분석 ✓					
	2 시트지	경 ☑ 첫 번째 두 번째 세 번째 직접지경 :			
텍스트마이닝		적용 할 <u>시트를</u> 지경합니다. 직접지경 : 예) 4.5.6			
매트릭스					

3 컬럼지정

감성분석

토픽분석

교 시각화

11

저장(S) 취소

인코딩(E): UTF-8

UTF-16 LE UTF-16 BE JTF-8 JTF-8(BOM)

- ※ 구축이 완료되면 수집하기에서 제공하는 기존 채널과 동일한 방법으로 이용하시면 됩니다
- ※ 수집된 데이터는 xlsx 파일 또는 txt 파일로 전달됩니다

- TEXTOM 소개 | 블로그 | 공지사항 | 문의하기 | 매뉴얼 | 관리자
- ※ 요청채널은 <u>수집된 데이터만 전달받는 방법(일회성)</u>과 텍스톰 요청채널 <u>페이지 내 구축하는 방법</u>이 있습니다

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

● 텍스톰이 제공하는 채널 외의 데이터 수집을 원하면 요청채널로 등록해주세요

IEXIUM			
🛢 수집하기	포털/SNS 뉴스 보유데이터 요청채널	1 요청채널추가	
		채널명	예) 뽐뿌게시판
에이터전처리 🗸	제목 작성된 제목으로 데이터가 저장됩니다.	URL	예) http://www.ppomppu.co.kr/zboard/zboard.g
수집리스트		토메인	예) ppomppu.co.kr
정제/형태소분석 = 데이터분석	기간 🛗 2020-04-17 ~ 🗰 2020-04-24 1주 1개월 1년 2220-04-24		요청하고자 하는 내용을 상세하게 입력하세요. 예) 1. 수집내용: 게시판의 제목, 본문, 날짜경보 2. 수집기간: 2017.01.01 ~ 2017.12.31
대트릭스 감성분석	^{채널} 사이트명 이 가지 이 가지 게시판명 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이 이	상세설명	
토픽분석 ····································	D KERMAN DEPUN		
시각화결과		췌다 기누에 대하	변드와 비용이 반생디며 했더이 특서에 따라 서적 기능이 번격
커스터마이징	C NEARCH-FE	될 수도 있습니다. 이 초기화	<u>2++ 4001 20</u> +41 412+ 4001 44 20 7104 100 ✓ 5록
	HIHRA DRATE DRATE DRATE DRATE DRATE		

요청채널

※ 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 <u>수집단위 기능</u>을 사용해주세요

- 전체수집으로 수집한 데이터는 키워드에 (전체 수집 결과)라고 되어있습니다
- ❷ 데이터미리보기에서 용량의 수치를 클릭하시면 수집된 데이터를 팝업창으로 미리 살펴볼 수 있습니다.
- <u>수집중에서 수집완료</u>까지 걸리는 시간은 설정해주신 수집 옵션에 따라 최소 10분에서 최대 5일이 소요됩니다



수집리스트 살펴보기

2. 데이터전처리

13

습니다

가 펼쳐집니다 ❷ 나뉘어진 데이터리스트를 개별적으로 선택하시어 데이터미리보기에서 <u>용량의 수치를</u> 클릭하시면 수집된 데이터를 팝업창으로 미리 살펴볼 수 있

● 수집단위 기능을 사용한 데이터는 키워드 앞에 폴더 아이콘이 있으며 해당 리스트를 클릭하면 수집 시 선택한 수집단위로 나뉘어진 하위 리스트

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 수집됩

※ 수집단위 기능이 뭔가요? 일/주/월/년의 시간 단위 중 선택된 단위로 데이터를 최대

ΤΕΧΤΟΜ	E	: 4	-집완료							뎈	<u>이터 용량추가</u> 더아이	엠씨님 ~ 🙎
🛢 수집하기		키워드검색 Q 검색결과 15506 / 15506 > · · · · ·							데이터미리보 용량을 클릭하면 *	기 해당 섹션에서 수집된	데이터 원문을 미리볼 수	≥ 있습니다.
▲ 데이터전처리 🛛 👻		T F 1 (01) 0							▶ 코로나 +메 2020-03-1	<u>르스</u> 3 ~ 2020-03-20		
수집리스트		포털/SNS	뉴스 보유네이터 .	요정재널					2020 03 1	5 2020 05 20		스지랴 시가하
정제/형태소분석	4	수집완료 리	스트의 데이터는 30일이 지나면 삭제	에됩니다.				Î				- 10 시 의
			키워드	채널	기간	수집날짜	용량	수집상태	채널	섹션	수집량 (건)	용량
∷ 데이터분석 ✓		0	100-00 - 000 W	STREET, STOPPED THE ST	2020/07/07 - 2020/07/04	200403.00	1.04.000	+0.00	-	블로그	7250	2. <u>39 MB</u>
텍스트마이닝			늘 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	네이버 NAVER	뉴스 카페	2042	<u>847 KB</u>
매트릭스			2020-03-20 ~ 2020-03-20				213 KB	스지와근	<u>[.11]</u>	지식인	307	<u>165 KB</u>
감성분석			2020 03 20 2020 03 20				215 10	+ 10		블로그	3541	1.17 MB
토픽분석			2020-03-19 ~ 2020-03-19				841 KB	수십완료	u≊ ⊳dłm	뉴스	2565	<u>866 KB</u>
교 시가하			2020-03-18 ~ 2020-03-18				909 KB	수집완료	Lat	카페	608	216 KB
삐 지극되 🔹			2020-03-17 ~ 2020-03-17				965 KB	수집완료	구글			
시각화결과			2020-03-16 ~ 2020-03-16				1006 KB	수집완료	Google	뉴스	310	<u>107 KB</u>
커스터마이징			2020-03-15 ~ 2020-03-15				688 KB	수집완료			ഥ 채널별	수집량 시각화

니다

수집단위 사용한 수집리스트 살펴보기

9

2. 데이터전처리

수집단위 사용한 수집리스트 살펴보기

техтом ≡

수집완료

<u>데이터 용량추가</u> 더아이엠씨님 🗸 👤



수집단위 기능을 사용한 데이터는 데이터미리보기에서 <u>기간별, 채널별 수집량 시각화</u>를 확인할 수 있습니다

2. 데이터전처리



수집된 데이터 미리보기

	데이터데	기리보기	
데이터명		생성날짜	용량
코로나 +메르스		2020-04-16	7.73 MB
네이버 다음 구글			
블로그 카페 뉴스 지식인			
2020.3.13.부동산 경제 뉴스 브리핑 realestate.daum.net 강남 3구 하락 이어졌지만9 불구하고 풍선효과로 https://blog.naver.com/jayreits?Redirect=Log&d	억원 미만 서울 외곽 ogNo=221851157	·경기 풍선효과 지속 신종코로나바이 068	▲ 러스감염증(코로나19) 확산에도
손 씻기의 중요성, 아무리 강조해도 지나치지 않아요 손 씻기만 잘해도코로나19 같은 많은 질병을 예방할 료기관은 물 m https://blog.naver.com/ohdamong?Redirect=Lo	<mark>と,</mark> 할 수 있어요. 제대로 ⁻ og&logNo=221852	손 씻는 지난메르스(급성 중동호흡 2197601	기증후군) 사태 이후, 전국의 의
<mark>당분간 매매일지는 쉬겠습니다.</mark> 코로나로 이렇게 망할 줄 몰랐습니다 그저메르스, 서 다 https://blog.naver.com/the816?Redirect=Log&	산정도로 생각했는 ogNo=221851375	데 참 사는게 막막해졌네요 갑자기 직 5750	안은 도움도 감사히 받겠습니
[책거울] 2020.3.12 거기다가 사스,메르스, 신종플루,코로나까지 5-6년 고 있는데요. 요즘코로나 https://blog.naver.com/studio_smallwave?Redi	주기로 바이러스 공 rect=Log&logNo=2	황에 시달리고 있고요. 어느덧 살아온 21851762978	: 시간의 절반이 공황도 함께 가

6)

- 데이터미리보기 팝업창 활용법 -

데이터미리보기 영역 <u>각 섹션의 용량을</u> 클릭하시면 왼쪽 이미지처럼 수집된 데이터를 팝업창으로 미리 살펴볼 수 있습니다

- 수집한 데이터는 채널과 그 채널의 섹션별로 분류되어 나타납니다
- ❷ 데이터미리보기 팝업창에서는 제목, 본문 요약, URL 을 제공하며, 각 데이터의 제목이나 URL을 클릭하시면 해당 데이터의 <u>실제 페이지</u>로 이동할 수 있습니다

❸ 데이터미리보기 팝업창으로, 수집된 데이터를 전체적으로 미리 살펴보며 Garbage Data(광고성 글, 원하는 주제와 무관 한 글)가 있는지, 수집이 정상적으로 잘 되었는지 체크해줍니다

Ж Тір

Garbage Data 는 정제/형태소분석 페이지에서 **키워드필터링** 기능으로 <u>제거 가능</u>하므로, 해당 문서들의 공통적인 키워드들 을 따로 메모해두시길 바랍니다

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

한 페이지에서 보여주는 데이터 수를 더 늘릴 수 있습니다

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

● 수집리스트에서 분석을 원하는 데이터를 선택합니다

을 수집하기	메르스	○ 김 경제/형태소분석 →	성색결과 19 / 18177			2 🔿	= 207H ✓ 107H 202H
🔺 데이터전처리 🔷 👻							2011
수집리스트	포털/SNS	뉴스 보유데이터	요청채널				
정제/형태소분석	수집완료 리	니스트의 데이터는 30일이 지나면 삭	제됩니다.				
		키워드	채널	기간	수집날짜	용량	수집상태
□ 데이디군식 ▼ 텍스트마이닝 □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □		2014/001	LENGERCONDUCTION AND COURTED AND WORK ORDERCONDUCTION AND REPORT FROM Proyect, ROLLING MICHINE AND ADDRESS OFF	2010/01/01 - 0220/07/00	200100-01	1.01.08	+288
메드닉스 	0	0.001400	1999 (2019)	20040-0-20040-0	20104-0	0.40	1000
토픽분석		늘 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
닚 시각화 ㆍ		2020-03-20 ~ 2020-03-20				700 KB	수집완료
시각화결과		2020-03-19 ~ 2020-03-19				1.32 MB	수집완료
커스터마이징		2020-03-18 ~ 2020-03-18				1.44 MB	수집완료
		2020-03-17 ~ 2020-03-17				1.48 MB	수집완료
		2020-03-16 ~ 2020-03-16				1.55 MB	수집완료
		2020-03-15 ~ 2020-03-15				1.11 MB	수집완료
		2020-03-14 ~ 2020-03-14				1.11 MB	스치이크



🗠 기간별 수집량 시각화

7950

2153

용량

5.01 MB

<u>1.7 MB</u>

수집량(건)

▶ 코로나 +메르스 2020-03-13 ~ 2020-03-20

채널

NAVER

용량을 클릭하면 해당 섹션에서 수집된 데이터 원문을 미리볼 수 있습니다.

데이터미리보기

섹션

블로그

뉴스

<u>데이터 용량추가</u> 더아이엠씨님 ~ 👤

수집리스트 선택하기

수집완료

Ē.

TEXTOM

2. 데이터전처리

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

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

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

디신시디 💙	직접선택	자동정제 선택안함	
리스트	분석목적에 맞는	세부적인 옵션을 지경할 수 있습니다.	▶ 코로나 +메르스
/혀태ㅅ브서			2020-03-20 ~ 2020-03-20
/ 8-11-2-7			분석리스트 데이터명을 지경하세요.
터분석 🗸	데이터정제		
ЕПЮІЦ			▶ 코로나 +메르스
리지	분리정제	 ● 전체(계목+본문) ○ 체목 ○ 본문 	2020-03-19 ~ 2020-03-19
 분석		제목과 본문을 분리하거나 통합하여 분석합니다.	분석리스트 데이터명을 지정하세요.
드 · 			
	키워드필터링	사용 사용안함 고초기학	▶코로나 +메르스
화 🗸		특경 키워드가 포함된 문서를 제거하거나 추출하여 경제/형태소분석 결과에 반영합니다.	2020-03-13 ~ 2020-03-13
화결과			분석리스트 데이터명을 지정하세요.
터마이징	중보제거	사요 사용아하	
	0111		리스트통합생성
			통합 생성할 분석리스트의 데이터명을 지정하세요.
	Window-Size	사용 사용안함	선택한 수집리스트를 통합하여 한 개의 <u>분석리스트</u> 로 생성합니다
		분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상 범위를 지정합니다. 주제어와 밀정하 다어들에 대하 데이터만 제공한니다	(<u>우십리스트</u> 는 용합되지 않습니다.)
			= 4

<u>데이터 용량추가</u> 더아이엠씨님 ~

, O 선택한수집리스트

전체선택

수집리스트 설정하기

정제/형태소분석

E

정제방법

TEXTOM

🛢 수집하기

▲ 데이

수집

정제

텍스

매트 감성 토픽

교 시각

시각:

ⅲ 데이

2. 데이터전처리

9

선택제외

<u>611 KB</u>

<u>707 KB</u>

<u>7.73 MB</u>

☰ 수집리스트

18

❸ 선택안함 이미 정제한 데이터를 업로드하여 분석리스트를 생성하고자 하는 경우 선택안함을 해주시면 됩니다

- ❷ 자동정제 텍스톰이 지정해놓은 설정값으로 자동 선택되어, 어느 정도 다듬어진 정제데이터를 볼 수 있습니다.
- **직접선택** 모든 설정을 이용자가 직접 선택할 수 있어 정교한 정제가 가능합니다

<u>수집한 데이터</u> 또는 <u>업로드한 보유데이터</u>의 정제, 형태소분석 방법을 지정할 수 있습니다

ΤΕΧΤΟΜ	글 정제/형태소분석	<u>데이터 용량추가</u>	ଗଦ୍ଦାଖ୍ୟାଧ୍ୟ 🗸 👤
을 수집하기	정제방법 · · · · · · · · · · · · · · · · · · ·	선택한수집리스트	
▲ 데이터전처리 →		전체선택	선택제외
수집리ㅅㅌ	분석목적에 맞는 세부적인 옵션을 지정할 수 있습니다.	▶ 코로나 +메르스	
지제/형태소부석		2020-03-20 ~ 2020-03-20	<u>611 KB</u>
		분석리스트 데이터명을 지정하세요.	
ⅲ 데이터분석 🔹	네이터성세		
텍스트마이닝		▶ 코로나 +메르스	
매트릭스	분리경제 (에 전체(제목+본문) 제목 본문	2020-03-19 ~ 2020-03-19	<u>707 KB</u>
감성분석	제목과 본문을 분리하거나 통합하여 분석합니다.	분석리스트 데이터명을 지정하세요.	
토픽분석			
133	키워드필터링 사용 사용안함 5 초기화	▶ 코로나 +메르스	
뇨 시각와 ◆	특경 키워드가 포함된 문서를 제거하거나 추출하여 경제/형태소분석 결과에 반영합니다.	2020-03-13 ~ 2020-03-13	7.73 MB
시각화결과		분석리스트 데이터명을 지정하세요.	
커스터마이징	중복제거 사용 사용안함		
		리스트통합생성	
		통합 생성할 분석리스트의 데이터명을 지정하세요.	
	Window-Size 사용 사용안함 분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상 범위를 지정합니다. 주제어와 밀접한 단어들에 대한 데이티만 제공합니다.	선택한 수집리스트를 통합하여 한 개의 <u>분석리스트</u> 로 생성 <u>(수집리스트</u> 는 통합되지 않습니다.)	성합니다.

정제방법 설정하기

2. 데이터전처리

19

데이터정제 설정하기

데이터정제 수집리스트 ▶ 코로나 +메르스 정제/형태소분석 2020-03-13 ~ 2020-03-13 7.73 MB 0 분리정제 전체(제목+본문) 🔿 제목 ○ 본문 분석리스트 데이터명을 지정하세요. ∷ 데이터분석 제목과 본문을 분리하거나 통합하여 분석합니다 텍스트마이닝 리스트통합생성 매트릭스 0 키워드필터링 사용안함 통합 생성할 분석리스트의 데이터명을 지정하세요. 감성분석 💿 제거 🛛 추출 선택한 수집리스트를 통합하여 한 개의 분석리스트로 생성합니다. 토픽분석 (수집리스트는 통합되지 않습니다.) 특정 키워드가 포함된 문서를 제거하거나 추출하여 정제/형태소분석 결과에 반영합니다. 圖 시각화 ☰ 수집리스트 시각화결과 6 중복제거 사용안함 커스터마이징 분석리스트생성 → OURL기반 🔵 내용기반 4 사용안함 Window-Size 키워드 사이즈 3개 \sim 분석하고자 하는 주제 키워드를 선정하고, 앞뒤 단어 개수를 기준으로 분석 대상 범위를 지정합니다. 주제어와 밀접한 단어들에 대한 데이터만 제공합니다.

● 분리정제 수집된 데이터에서 분석에 사용할 텍스트 영역을 지정할 수 있으며, 제목, 본문, 제목+본문 중에서 선택 가능합니다

⑦ 키워드필터링 특정 키워드가 포함된 문서들을 제거하거나 추출할 수 있습니다

※ <u>한 필드에 한 단어 씩</u> 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 2. 데이터전처리

데이터정제 설정하기

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

※ 내용기반은 특수문자, 띄어쓰기까지 완전히 동일할 때 제거됩니다

※ URL기반은 링크 전체가 완전히 동일할 때 제거됩니다

❸ 중복제거 중복된 데이터를 제거할 수 있습니다

수집리스트	데이터정제		
지지/형태소분석 3제/형태소분석 ■ 데이터분석		● 전체(제목+본문)	▶ 코로나 +메르스 2020-03-13 ~ 2020-03-13 Z.73.MB 분석리스트 데이터영을 지정하세요.
택스트마이닝 매트릭스 2 감성분석 토피분석	키워드필터링	사용 사용안함 그 초기화 • 제거 추출 키워드추가	리스트통합생성 통합 생성할 분석리스트의 데이터명을 지경하세요. 선택한 수집리스트를 통합하여 한 개의 <u>분석리스트</u> 로 생성합니다. (<u>수집리스트</u> 는 통합되지 않습니다.)
 ▲ 시각화 ✓ 사각화결과 거스터마이징 	중복제거	특정 키워드가 포함된 문서를 제거하거나 추출하여 정제/형태소분석 결과에 반영합니다. 사용 사용안함 ● URL기반 내용기반	= 수집리스트 분석리스트생성 →
()	Window-Size	사용 사용안함 키워드 사이즈 3개<	



- ❷ 분석기 각 분석기의 특성을 참고하여 형태소분석기를 선택해줍니다 Espresso K 와 MeCab 분석기 차이 자세히보기
- 분석언어 데이터를 인식할 언어를 선택해줍니다

주는 패키지를 의미합니다

형태소분석 설정하기

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

을 수집하기					^	서태하스지리스트	
	형태소분석				>	고레서타	서태제이
데이터전처리	부선어어	하구에 역에 주구에				전세선적	
수집리스트	문득간학					▶코로나 +메르스	
정제/형태소분석						2020-03-13 ~ 2020-03-20	<u>3.65 MB</u>
2	분석기 😮	Espresso K MeCab				분석리스트 데이터명을 지정하세요.	
데이터분석 🗸							
텍스트마이닝		예) "사회복지학과"를 Espresso K는 "사회복지학"으로 Mecab은 "사회 복지 학과"로 정제	합니다.			리스트통합생성	
매트릭스						통합 생성할 분석리스트의 데이터명을 지정하세요.	
감성분석	нити				Ľ.	선택한 수집리스트를 통합하여 한 개의 <u>분석리스트</u> 로 생성 (수집리스트는 통합되지 않습니다.)	영합니다.
토픽분석	군식품사	선산물사 경제품사 1 물사태그		_ 3 조기와			
							≡ 수집리스트
시각화 🗸		Y 물건 경지(WWG) Y 가파경지(WWF) Y 의근경지(WWG) Y 근처경					
시간하격과		▶ 용언	▶ 어근				
		농사(VV) 영용사(VA)	어근(XR)			분석리스트생성 →	
기스티마이징		▶ 수식언	▶ 독립언				
		관형사(MM) 일반부사(MAG) 접속부사(MAJ)	감탄사(IC)				
		▶ 접미사	▶ 접두사				
		명사접미사(XSN) 동사접미사(XSV) 형용사접미사(XSA)	체언접두사(XPN)				
		▶ 한글 이외					
		외국어(SL) 숫자(SN)					

2. 데이터전처리

- ④ 품사 태그 분석기별 품사 태그를 확인할 수 있습니다
- ❸ 결과 미리보기 선택한 품사가 예시문장에 바로 적용되어 정제데이터에 선별될 단어를 미리 확인할 수 있습니다
- ② 단순품사 <u>명사, 형용사, 동사, 외국어, 숫자</u>를 선택할 수 있습니다
- 상세품사 <u>체언, 용언, 어근, 수식언, 독립언, 접미사, 접두사, 외국어, 숫자</u>를 선택할 수 있습니다

분석품사

분석품사	단순품 1 상세품사 4 품사태그		3 초기화 결과 미리보기 (<u>수집리스트는</u>	프로플 등입하여 한 개의 <u>포격디스트</u> 로 868입니다. - 통합되지 않습니다.)
	▶ 체언 ✔ 일반 명사(NNG) ✔ 고유명사(NNP) ✔ 의존명사(NNB) ✔ 단	위명사(NNBC) 🔽 수사(NR) 🗹 대명사(N	(P)	≡ 수
Ť	▶ 용언 □ 도비AAA □ 전 원이 (AAA)	▶ 어근	·····································	결과 미리보기 (분석품사)
-	중사(VV)) 더근(XR) > 독립언] 감탄사(IC) 	· 체언 · 체언 · 월명 명사(INIG) · 고유명사(INIP) · 의존명사(INIS) · 용연 · 동네(AAA · · · · · · · · · · · · · · · · · ·	관속 중사 전역 문위병시(NHBC) 로 수시(ND) 로 대왕서(NP) · 여근 · 여근 · 이금/VP)
	▶ 접미사	▶ 접두사 체언접두사(XPN)	 > 수석연 국청사(MM) 일반부사(MAG) 감숙부사(MAJ) > 컵미사 정사컵미사(VSN) 중사컵미사(VSN) > 하금 otg 	 → 10,000 → 400 - 2004 - 2004
	외국어(SL) 숫자(SN)		의국아(SL) 숫자(SN) 에시문장	형태소 분석 문장
нити			서울역에서 시청 앞까지는 지하철 1호선 한 구간의 거리입니다.	서울역(NNP)+에서(JKB)+ 시청(NNG)+ 앞(NNG)+까지(XX)+는(JX)+지하 (NNG)+ 1(SN) +호(NNBC)+선(NNG)+ 한(MM)+ 구간(NNG)+의(JKG)+ (NNG)+ 입니다(VCP)
문석품사	▲ 단압품사 실세품사 ● 품사 태그 ✓ 명사 형용사 동사 외국어 숫자		아곳의 중간통에는 국보 1인 승례문이 있지만 안타깝게도 화마로 불타 지는 건속을 위해 가답마을 쉽지 해준 상태입니다.	○ 귀(N)-9(U(N)-62(N)-64(N)-64(N)-64(N)-64(N)-44(N)-64(N)
			그러면 함께 서울역에서 시청 앞까지 유유자격하게 걸어 볼까요? LET'S GC	그러면(MAJ)+함께(MAG)+ 서울억(NNP) +에서(JKB)+ 시청(NNG)+앞 (NNG) +까지(JX)+ 유유저척(NNG) +하(XSV)+게(EO)+Ż(VV)+어(EO)+볼 (VX)+LET(SL)+'(SY)+S(SL)+GO(SL)+!(SF)
			짜진 이곳이 서울 4대문 중 하나인 숙경문입니다. 우속을 보시면 치마 옆 쪽 소나무의 자태가 고봉스럽습니다.	재진(이는 이곳(NP) + 이(NS)+서울(NNP)+4(SN)+대물(NNG)+중(NNB) (ND)-인(NCP)+도NN+중성ዊ(NNG)-9U(LFN(NCP)+우운(NNG)+2(N (N)+4(D)-10(NG)+2(NNG)+2(NNG)+2(N)+2(D) (LFQ)NG)+2(NG)+4(NNG)+7(NS)+28(NNG)+2합(SA)+습 (LFQ)NG)+2(NG)+4(NNG)+7(NS)+28(NNG)+2합(SA)+습

형태소분석 설정하기

주시면 됩니다

단어 사용여부도 매번 설정 가능합니다

3 그룹지정 그룹지정 란에서 사용할 사전의 그룹명을 선택해

❷ 사용자사전설정 그룹별로 단어를 등록할 수 있으며, 등록된

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

 ▶ 한글 이외 □ 외국어(SL) □ 숫자(SN) 	+ 새그
2 1 사용자사전 ②	
3 그룹지경 □지경 ✓	
경제할 키워드를 사용자사전에 먼저 등록해주세요. (마이페이자-사용자사전) 텍스톰을 이용하여 여러 번 분석하실 경우, 사용자 사전을 이용하시면 반복적인 작업 없이 빠르고 효율적인 분석이 가능합니다. 예)'워라벨'을 '워라밸'로 변경	
+ 요청채널추가	탄어검쓰 흡 삭제

+ 새	그룹만들기 💼 삭제		* ^	* *
		그룹명		^
~	미지정			
	***)			
	842			
	8917-064			
-				-
단어검	색 Q 검색결과	553671 / 553671	=	10개 🗸
前 삭지	세 사용 미사용 다운로드(엑셀)	엑셀일괄등록	전	स
	はさきにの		на	
	면경일단어	→ 주성년에	면성	
"변경할[Tip. 변경	단어"가 아래 리스트에 중복으로 등록된 경우에 1할단어의 중복 여부는 상단 '단어검색'을 통해	는 최근에 등록된 내용으로 적용됩니다. 확인할 수 있습니다.		
	변경할단어	수정단어	등록일	사용여부
	인사	인사정책	2020-05-03 12:09:17	Ο
	트랜드	트렌드	2020-05-01 15:44:35	
	주시꾸띠르	쥬시꾸뛰르	2020-05-01 15:44:15	
	소매부	소매부분	2020-05-01 02:02:03	
	전보	전보다	2020-05-01 02:00:16	
	예전	예전에	2020-05-01 01:59:55	
	전	전에	2020-05-01 01:58:05	
	이번	이번에	2020-05-01 01:57:46	
	에슬레저	애슬레저	2020-04-30 18:47:15	
	수 있는원피스	수 있는 원피스	2020-04-30 17:08:52	

사용자사전 설정하기



❸ 데이터 편집 고유명사 및 복합명사, 동의어, 불용어 제거 등 데이터 수정을 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.

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

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

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

ΤΕΧΤΟΜ	들 텍스	노트마이닝					<u>데이터 용량추</u> 2	I 더아이엠씨님 ~ 🙁	
을 수집하기	메르스	Q 검색결과 23 / 21454			≡ 107# ~ >	텍스트마이닝 매	트릭스 감성분석	토픽분석	
▲ 데이터전처리 ✓	포털/SNS	뉴스 보유데이터 요청채널	형태소 분석이 완료되면, 바로편집하기/업로드를 통해 단어를 정제해보세 웹 상에서 빠르고 쉽게 단어 편집을 하고자 할 경우에는 바로편집하기 가능 정계 데이터를 내려 받아 작업을 하고자 할 경우 업로드 기능을 사용하세요 위무데이터						
정제/형태소분석		데이터명		생성날짜	용량	미리보기	다운로드(Excel)	다운로드(txt)	
≡ 데이터분석 🛛 🗸		코로나 +메르스	ľ	2020-03-20	6.26 MB	정제데이터 2			
		코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13		2020-03-20	6.26 MB	미리보기	다운로드(Excel)	다운로드(txt)	
넥스드마이닝		1944 - Contra 19	127	00940-04	201.01				
매트릭스		1944	15	20140-01	1.21.00	데이터 편집 😗	편집 된 더	이터가 적용되어 있습니다.	
감성분석		194A	12	00040.00	121.00		바로편집하기		
토픽분석		644.008	15	00040.04	20.00.00	바로편집하기 별도의 다운로드 없이, 원	빈상에서 데이터를 바로 편집	할 수 있습니다	
시가하 🗸		104-51	17	20040-01	20.00.000			· · · · · · · · · · · · · · · · · · ·	
II 시극의 ·		194A	12	2210-0214	3.0	파일업로드		90° •	
시각화결과		6a.	12	010110	377.03	비 = 입국그 원문데이터가 아닌 정계데이터를 다운로드하여 단어 편집을 진행 후 ?			
커스터마이징		6a.	완료된 데이터를 업로드 합니다. - 엑셀 파일 형식의 데이터를 txt 파일로 반					로 인코딩)하여 단어편집	
						후 업로드합니다. -'편집된 데이터가 적용5 사용할 수 있습니다.	티어 있습니다'라는 텍스트가	뜨면 파일 업로드 기능을	

텍스트마이닝하기



- **부분일치**는 '이번v메르스v사태', '이번메르스사태'를 '이번v코로나v사태', '이번코로나사태'로 '메르스'를 포함한 **모든 문자열을 변경**합니다.

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

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

부분일치는 문서에서 부분적으로 일치하는 문자열을 포함하여 모두 변경합니다.

❸ 정확한일치/부분일치 정확한일치는 정확하게 일치하는 문자열만 변경합니다.

ТЕХТОМ	글 텍스트마이닝	정제데이터 (텍스트마이닝)	데이터 용량추가 더아이엠씨님 ~ 오
 ➡ 수집하기 ▲ 데이터전처리 ✓ ▲ 수집리스트 정제/형태소분석 	메르스 Q 검색결과 23 / 클 삭제 포털/SNS 뉴스 보유데이터 요청채널	데이터명 생성날짜 용량 코로나 + 메르스 2020-05-11 6.26 MB 응 ✓ 변경한 내용 적용 미리보기 바로편집하기 3 이 분분일치 6	택스트마이닝 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 미리보기 다운로드(Excel) 다운로드(txt)
	코로나 +메르스 코로나 +메르스 [수집단위] 2020-03-1	④ 변경할단어 + → 수정단어 변경 사용자사건 수경은 사용자사건 메뉴에서 가능합니다. 사용자사건 바로가기 >	MB MB 미리보기 다운로드(Excel) 다운로드(txt)
배트릭스 감성분석 토픽분석		6 2 5 추천단이 1 고로나 검사 후기 고로나검사 왕졸보관 선별진료소 수정내역 고로 사소 메르스 일 남일 내가고로나검사 도로 사소 메르스 입 남일 내가고로나검사 * 천진자 확진자 12 특집 기사 코로나19 현황 전망 중증 사망 수 자료 따 확진자	데이터 편집 편집 된 데이터가 적용되어 있습니다. 1 바로편집하기 바로편집하기 별도의 다운로드 없이, 웹상에서 데이터를 바로 편집할 수 있습니다
교 시각화 · · · · · · · · · · · · · · · · · · ·		시료 효과 11 서계 검체법 워킹스루 진료소 귀감 드라이브스루 워 지료 효과 11 킹스투 도입 메르스때 이어코로나바이러스 초기 검사 김료 효과 비이러스 진실 이제갑 교수 바이러스 경우 중환자 치 료 대의 치사율 메르스 달리코로 무한 우한 제외 도 신종플루 10 신종플루 자발 자가격리 달 육아 일상 활동 15년 이번코로나19 전 세계 국결 차단 신종 플 외환위기 메르스 등 일	 □ ○ <li< td=""></li<>
		전 세계 국경 차단 신종 풀 외환위기 메르스 등 일 ▼ 이번	-편집된 데이터가 적용되어 있습니다'라는 텍스트가 뜨면 파일 업로드 기능을 사용할 수 있습니다.

텍스트마이닝하기

※ 변경한 내용 적용 버튼을 누른 후에는 되돌릴	수 없습니다
----------------------------	--------

텍스트마이닝하기

● 수정내역 되돌리기 아이콘을 클릭하면 해당 단어의 변경을 취소할 수 있습니다.

※ 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.

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



텍스트마이닝하기



※ 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** <mark>있습니다.</mark>

❷ 편집한 데이터의 적용 진행 상황을 확인할 수 있습니다.

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

※ 원문데이터가 아닌 정제데이터를 다운로드 받아 편집 해야 합니다.

● **파일업로드** 정제데이터를 다운받아 단어 편집 후, 편집이 완료 된 파일을 업로드 하는 데이<mark>터 편집 방법입니다.</mark>

ΤΕΧΤΟΜ	로 텍스	스트마이닝					<u>데이터 용량추</u> 2	1 더아이엠씨님 v 😍	
 ● 수집하기 ▲ 데이터전처리 ◆ 수집리스트 	메르스	Q 검색결과 23 / 21454 뉴스 보유데이터 요청채널			≡ 107₩ ✔)	텍스트마이닝 매 형태소 분석이 완료되면 웹 상에서 빠르고 쉽게 당 정체 데이터를 내려 받아	트릭스 감성분석 , 바로편집하기/업로드를 통 단어 편집을 하고자 할 경우이 ት 작업을 하고자 할 경우 입5	토픽분석 해 단어를 경제해보세요. 네는 바로편집하기 기능을, 르드 기능을 사용하세요.	
정제/형태소분석		데이터며		새서날짜	요랴	원문데이터			
늘 데이터분석 ◆		코로나 +메르스	ľ	2020-03-20	6.26 MB	3 전계데이터		이군도=(00)	
		코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13	ľ	2020-03-20	6.26 MB	미리보기	다우로드(Excel)	다우로드(tyt)	
텍스트바이닝		1944 -	10	2009-02-04	201.00	-101-221		-12(00)	
매트릭스		64.5	15	2004-02-01	17100	데이터 편집	2 편집 된 더	비이터가 적용되어 있습니다.	
감성분석		64.5	18	0040-0	12100		바로편집하기		
토픽분석		644.008	15	222742-04	2.0.10	바로편집하기 별도의 다운로드 없이, ⁶	볩상에서 데이터를 바로 편집	l할 수 있습니다	
교 시가하 🗸		64-59	15	2009-02-03	2,2,2,1	•		▲ 전요	
		6a.	15	2210-02-04	1.0	파일업로드			
시각화결과		1944 -	15	010110	27.0	원문데이터가 아닌 정제	데이터를 다운로드하여 단어	비편집을 진행 후 정제가	
커스터마이징		- (int.	17	29112	3,9.00	- 전보선 네이더글 접도는 접니다. - 액셜 파일 형식의 데이터를 txt 파일로 변경(UTF-8로 인코딩)하여 단어편집 - 후 업로드합니다.			
					•	-'편집된 데이터가 적용!	티어 있습니다'라는 텍스트가	뜨면 파일 업로드 기능을	

텍스트마이닝하기

❷ N-gram n개 단어의 연쇄를 의미합니다.

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

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

● 단어빈도 추출된 단어와 데이터 내 해당 단어의 빈도수를 나타냅니다.

	데이터	B		생성날짜		용량			▲ ✓ 적용
코로나 +메르스 🗹			2020-03-20)	6.26 MB	파일업로드 위문데이터가 아닌 정제데이터를 다우로드하여 다어 펴진을 지행 흐 정제기			
	분석결과 (텍스	<u>스트</u> 마이닝)		2020-03-20)	6.26 MB	완료된 데이터를 업로드 - 엑섹 파익 형식이 데이	합니다. 터르 txt 파익로 벼경(LITE-8	(루 이코딩)하여 다어펴?
데	이터명	생성날짜 2020-05-13	용량 6.26 MB	2	분석결과 (텍스트마이닝)		후 업로드합니다. - '편집된 데이터가 적용!	티어 있습니다'라는 텍스트가	· 뜨면 파일 업로드 기능을
단여빈도 N-gram TF-IDF	연결중심성			데이터명	생성날짜	88	사용할 수 있습니다.		
·····································	니다. 전체 단어는 다운로드하여 확인할 수 있	습니다.	ц <u>я</u>	코로나 +메르스	2020-05-13	6.26 MB	비서경기		
단어	빈도	백분율(%)	누적백분율(%)	단어빈도 N-gram TF-IDF 연결중심성			문역열파		
메르스	296	2.661	2.	상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다	운로드하여 확인할 수 있습니다.	다운로드	▶ 단어빈도 🕕		
코로나	277	2.491	5.	단어1	단어2	빈도	미리보기	다운로드(Excel)	다운로드(txt)
사스	248	2.23	7.	코로나	바이러스	101			
바이러스	238	2.14	9.	코로나	19	96	►N-gram 🙎		
것	129	1.16	10.	사스	메르스	73	미리보기	다운로드(Excel)	다운로드(txt)
19	103	0.926	11.	신종	코로나	42			
개반	92	0.827	12.	메르스	재즈	39	→ TF-IDF 3		
소	73	0.656	13.	신종플루	" 메르스	21	미리보기	다우르드(Excol)	다우ㄹㄷ(+++)
둡	71	0.638	14.	2015년	메르스	19			이근포_(00)
신종플루	65	0.584	15.	백신	개발	19	▶ 연결중심성 4		
사태	63	0.566	15.	문재인	대통령	19			
신종	60	0.539	16.	메르스	사태	18	미리모기	나운도느(Excel)	나군도느(txt)
백신	60	0.539	16.	사스	비교	17	▶ 개체명인식		
치료계	60	0.539	17.	메르스	5	17			
				치료제	개발	16		미리보기	
			4 2 2	바이러스	감염증	15 💌			

분석결과 → 단어빈도 1		
미리보기	다운로드(Excel)	다운로드 <mark>(txt)</mark>
►N-gram		
미리보기	다운로드(Excel)	다운로드(txt)
► TF-IDF		
미리보기	다운로드(Excel)	다운로드(txt)
• 연결중심성		
미리보기	다운로드(Excel)	다운로드(txt)
▶ 개체명인식 5		
	미리보기	

텍스트마이닝하기

※ TF – IDF = TF X 1/DF ※ 특정 범위 내에서 모든 단어들의 빈도수와 단어가 포함된 문서(정보)들의 빈도수를 구한 후 역수를 취해 곱하여 문서의 중요도를 찾는 방법

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

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

	(텔) 분석결과 (텍스	트마이닝)		용량			▲
코로나 +메르스	데이터명	생성날짜	용량	6.26 MB	파일업로드 원문데이터가 아닌 경제[와르되 데이터르 언르드	헤이터를 다운로드하여 단어 하니다	편집을 진행 후 정제가
코로나 +메르스 [수집단위] 202(메르스	코로나 +메르스 다이비드 N-oram TEHDE 여건증심성	2020-05-13	6.26 MB	6.26 MB 300 KB	전묘된 데이더들 접도드 합니다. - 엑셀 파일 형식의 데이터를 txt 파일로 변경(UTF-8로 인코딩)하여 단어편경 후 업로드합니다. -편집된 데이터가 적용되어 있습니다'라는 텍스트가 뜨면 파일 업로드 기능을 사용할 수 있습니다.		
메르스	상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습	니다.	다운로드	2.77 MB			
메르스	단어		283 250320606	2.77 MB	분선격과		
메르스 감염병	코로나		253.023732955	30.05 MB	▶ 단어빈도 1		
메르스	사스		224.194243663	9 KB	미리보기	다운로드(Excel)	다운로드(txt)
메르스	것 		200.694460474	377 KB	►N-gram 2		
메르스	19		181.845628868	5.95 MB	미리보기	다운로드(Excel)	다운로드(txt)
	때		175.590499516		► TF-IDF		
	Υ. 		153.768523445		미리보기	다운로드(Excel)	다운로드(txt)
	5		145.251339037		▶ 연결중심성 4		
	시프에		144.585585849		미리보기	다운로드(Excel)	다운로드(txt)
	신종플루		143.315783142		▶ 개체명인식 5		
	2015년		134.069611302 💌			미리보기	

텍스트마이닝하기

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

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

⑦ 개체명인식 단어를 13개의 개체명 범주에 따라 분류하고 그 빈도수를 나타냅니다.

④ 연결중심성 A라는 단어가 B라는 단어와 얼마나 많이 연결되어 있는지를 나타냅니다.

	데이터명		생성날짜	용량			▲ ✓ 적용
코로나 +메르스			2020-03-20	6.26 MB	파일업로드 워문데이터가 아닌 경제	데이터를 다운로드하여 단어	편집을 지행 후 경제가
4 분석결과 (텍스트마아)	닝)		2020-03-20	6.26 MB	완료된 데이터를 업로드	: 합니다.	
데이터액	새선날짜 유타				- 엑셀 파일 형식의 데이 후 업로드합니다.	터를 txt 파일로 변경(UTF-8	로 인코닝)하여 난어편십
코로나+메르스	2020-05-13 6.26 MB				·'편집된 데이터가 적용!	되어 있습니다'라는 텍스트가	뜨면 파일 업로드 기능을
단어빈도 N-gram TF-IDF 연결중심성		사람 학	문 대상물 기관 지역 문명 날짜 시간 숫자	· 사건/사고 동물 식물 금속 용어	사용할 수 있습니다.		
상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.	다운 <u>로드</u>	※ 다운로드시	전체 문서를 확인할 수 있습니다.	다운로드	분석결과		
단어	연결중심성		Word	Frequency	• 다어히는 🕕		
바이러스	0.0563200815494		Ż	54			
메르스	0.0519877675841		전모두	5	미리보기	다운로드(Excel)	다운로드(txt)
	0.0491845056065		후보물질의코	5			
사스	0.046381243629		키트	4	▶ N-gram		
코로나	0.0445973496432		박근혜	4	미리보기	다운로드(Excel)	다운로드(txt)
<u></u>	0.025993883792		김	3			
5	0.0237003058104		폭	3	→ TF-IDF 3		
сł	0.0234454638124		곽승준	3			510.754.0
19	0.0226809378186		박영선	3	미리모기	나군도느(Excel)	나군도느(txt)
개발	0.0203873598369		윤영호	2	▶ 여겨주신성 4		
사태	0.0175840978593		조셉김	2			
감염	0.0175840978593		이재갑	2	미리보기	다운로드(<mark>Excel</mark>)	다운로드(txt)
한국	0.0155453618756		유동우	2			
감염병	0.0155453618756		이재명	2	▶ 개제명인식 🕤		
백신	0.0150356778797 👻		이낙연	2 -		미리보기	
	1 2						

텍스트마이닝하기

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

※ 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

수학적 분석을 위한 소셜네트워크 데이터의 기본 형태는 **테이블 형태의 매트릭스**입니다.

정제된 단어들 간의 동시출현빈도 (공출현빈도: co-occurrence)로 작성된 매트릭스 데이터를 제공합니다.

						\bullet
ТЕХТОМ	글 매트릭스					<u>데이터 용량추가</u> 더아이엠씨님 ~ 👤
● 수집하기 ▲ 데이터전처리 ×	메르스 Q 검색결과 23 / 21556 효 삭계			≡ 107∦ ✓ >	텍스트마이닝 매트릭스	감성분석 토픽분석
수집리스트	포털/SNS 뉴스 보유데이터 요청채널			데이터적용 리스트	분석에 활용할 단어를 선택해보세요 싶은 단어에 대한 결과만을 확인할	바로선택하기/업로드를 통해 분석하고 수 있습니다.
정제/형태소분석	데이터명		생성날짜	용량	1-mode	2-mode
∷ 데이터분석 ✓	코로나 +메르스	Ľ	2020-03-20	6.26 MB	Hł	로선택하기
	코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13	Ľ	2020-03-20	6.26 MB		📱 🗸 적용
텍스트바이닝	메르스	Ľ	2020-02-24	300 KB	🖥 예시파일 다운로드	
배트릭스	메르스	Ľ	2020-02-20	2.77 MB	추출한 단어를 txt 파일/Excel I	ት일(UTF-8로 인코딩)로 업로드 해보세
감성분석	메르스	Ľ	2020-02-20	2.77 MB	<u>.</u>	
토픽분석	메르스 감염병	Ľ	2020-02-24	30.05 MB		
교 시가하는 것	메르스 질병	Ľ	2020-02-21	58.95 MB	분석단어	
	메르스	ľ	2018-12-14	9 KB	▶ 선택단어	미리보기 다운로드
시각화결과	메르스	ľ	2018-11-29	377 KB		
커스터마이징	메르스	ľ	2018-11-29	5.95 MB	분석결과	

매트릭스 만들기



3. 데이터분석

텍스트마이닝 매트릭스	: 감성분석	토픽분석				
분석단어선택 분석에 활용할 단어를 선택해보세요. 바로선택하기/업로드를 통해 분석하고 싶은						
단어에 대한 결과만을 확인할 4	수 있습니다.	2-mode				
- Induc		2 11000				
2	단어	선택이 적용 되어있습니다.				
	마도신댁야기	▶ ✓ 적용				
■ 예시파일 다운로드 추출한 단어를 txt 파일/Ex	cel 파일(UTF-8로 인	코딩)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				

분석하고 싶은 단어를 웹상에서 바로 선택하거나, 딘						
_	분석	단이선택 (1-mode)				
	데이터명 국리나니메르스	생성 2020년	날짜	용량 6.26 MR		
	단어빈도		TF-IDF	0.20110		
▶ 선택단어	범수 확인 → 선택단어누격비율	%		다운로드 적용		
상위 2007	개 까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 ? 	있습니다. 	백분율 (%)	누적비율(%)		
	메르스	296	2.661%	2.661%		
	코로나	277	2.491%	5.152%		
	사스	248	2.23%	7.382%		
	바이러스	238	2.14%	9.522%		
	2	129	1.16%	10.682%		
	19	103	0.926%	11.608%		
	ų	92	0.827%	12.435%		
	개발	81	0.728%	13.163%		
	수	73	0.656%	13.819%		
	5	71	0.638%	14.458%		
	신종플루	65	0.584%	15.042%		
	АŀЕł	63	0.566%	15.609%		
	신종	60	0.539%	16.148%		
	백신	60	0.539%	16.688%		
	치료제	60	0.539%	17.227%		
	2015년	56	0.504%	17.731%		
	감염	52	0.468%	18.198%		
	비교	47	0.423%	18.621%		
	감염병	47	0.423%	19.043%		

● 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.

바로선택하기

아이리스트를 직접 업로드 할 수 있습니다.

 분석단어
 성 선택한 단어와 빈도를 미리보기 또는 다운로드 받을 수 있습니다.

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

※ 작성한 매트릭스로 UCINET를 활용하
여 시각화를 하실 수 있습니다.

UCINET 활용법 알아보기

분석단어선택 (1-mode)



● 단어빈도 또는 TF-IDF 순으로 단어를 볼 수 있습니다.

🕗 선택 단어 수

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

※ 가시성이 좋은 시각화를 만들기 위하여 선택 단어 수를 <u>50~70개</u> 사이로 선택 할 것 을 추천 드립니다.

단어선택

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

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

상위	200개 까지	단어를 미려	리볼 수	: 있습니다.	전체 [단어는 !	다운로드하여	확인할 수	- 있습니다.

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



	A	В	C	D	E	F
A	0	1	1	0	0	0
В	0	0	0	2	0	0
С	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

※ 1-mode와 2-mode의 차이점은 무엇인가요?

- 1-mode 의 경우, 매트릭스와 같이 행과 열이 같은 단어로 이루어진 매트릭스 (모든 네트워크의 가장 일반적인 유형)
- 2-mode 의 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, 동호회화 회원, 구직자와 취업박람회, 영화와 영화배우 간의 관계)

<u>2-mode로 분석한 사회 이슈 자세히 보기</u>



	Sports	Movie	Cook
A	1	1	0
В	0	1	0
С	1	0	1
D	0	0	1
E	0	1	1
F	1	0	1

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

TEXTOM	글 매트릭스						데이터용량추가 더아이엠씨님 ~ 👤
	2 정제단어 빈도						
S 7 ⁴ °	단어	빈도			≡ 10 ⁷ स ∨ >	텍스트마이닝 매트릭스	감성분석 토픽분석
	메르스	296	1				
▲ 데이터저처리	코로나	277				분석단어선택	
	사스	248			데이터져요 리스트	분석에 활용할 단어를 선택해보	세요. 바로선택하기/업로드를 통해 분석하고
수집리스트	바이러스	238				싶은 단어에 대한 결과만을 확인	할 수 있습니다.
	것	129				1-mode	2-mode
성세/영태소문	19	103		생성날짜	용량		
	αų	92	-1				단어 선택이 적용 되어있습니다.
∷ 데이터분석	개발	81		2020-03-20	6.26 MB		바르서태하기
	<u></u>	73		2020-03-20	6.26 MB		미도전국하기
텍스트마이닝	5	71		2020-02-24	300 KB		🖹 🗸 적용
매트린스	신종플루	65		2020 02 24	47.000		
	사태	63		2	2.77 MB	■ 예시파일 다운로드	
감성분석	신종	60		2020-02-20	2.77 MB	주출한 난어들 txt 파일/Exc	el 파일(UTF-8로 인코딩)로 업로드 해보세
토피부선	백신	60	_			ш.	
	치료제	60		2020-02-24	30.05 MB		
	2015년	56		2020-02-21	J. HB		
표 시각화	김염	52		2018-12-14	9 KB	문석단어	
시각화결과	메르스			2018-11-29	377 KB	▶ 선택단어	미리보기 다운로드

5

분석결과				0 5	<u></u> 년	빈도	드의	매	Ē	릭스	<u> </u>												
→ 단어빈도	미리보기	다운로드	LE							분	넉결과	(1-ma	ode) <mark>-</mark>	빈도									
▶ 엣지리스트	미리보기	다운로드		E a	코로 다	사스	바이러스	것	19	때 갤	Ŷ	Цö	-Truck to by -	사태	시공	백신	친 2 료 제	015 년	가이면이	비교	감염병	하고	· · · · · · · · · · · · · · · · · · ·
				메르스) 207	229	168	83	65	74 36	36	50	51	52	38	40	37	44	25	40	27	22	25
▶ 유클리디언계수	미리보기	다운로드		코로 나 2	07 0	176	230	77	147	42 36	i 40	22	55	44	64	39	10	52	29	26	12	30	43
				사스 2	29 176	0	183	87	45	58 20	23	39	36	17	50	32	20	42	23	34	26	13	31
▶ 고자인세수	미리보기	다운로드		바이 10 러스 1	58 230	183	0	71	34	19 35	32	24	18	12	77	21	35	28	49	12	9	14	26
				것 8	3 77	87	71	0	27	30 6	15	8	12	13	22	12	11	25	8	9	5	6	17
› 사카느계수	미리보기	다운로드		19 6	5 147	45	34	27	0	22 11	25	7	31	16	7	9	3	17	7	9	13	7	10
				때 7	4 42	58	19	30	22	0 0	19	7	20	24	6	2	3	10	7	16	10	9	8
▶ 상관계수	미리보기	다운로드		개발 3	6 36	20	35	6	11	0 0	15	9	16	8	10	56	45	2	9	0	14	6	7
				수 3	6 40	23	32	15	25	19 15	0	7	17	6	6	10	12	2	8 O	겠기	(2)	12	6
				등 5	0 22	39	24	8	7	7 9	7	0	15	11	7	7	4	5	5	6	22	2	3

20

16 17 15



2

단어빈도

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

❷ 엣지리스트(Edge List)

단어와 단어, 노드와 노드를 짝지어 목록으로 보여줍니다.

분석결과 (1-mode) - 엣지리스트								
word1	word2	weight						
메르스	코로나	207						
메르스	사스	229						
메르스	바이러스	168						
메르스	것	83						
메르스	19	65						
메르스	αң	74						
메르스	개발	36						
메르스	수	36						
메르스	듕	50						
메르스	신종플루	51						
메르스	사태	52						
메르스	신종	38						
메르스	백신	40						
메르스	치료제	37						
메르스	2015년	44						
메르스	감염	25						
메르스	비교	40						
		Ŧ						

단어 간의 유사도 측정방법 자세히 알아보기

- 상관계수
- ❸ 자카드 계수
- ❷ 코사인계수
- 유클리디언계수

분석결과		
▶ 단어빈도	미리보기	다운로드
▶ 엣지리스트	미리보기	다운로드
• 유클리디언계수	미리보기	다운로드
2 > 코사인계수	미리보기	다운로드
3 › 자카드계수	미리보기	다운로드
4 › 상관계수	미리보기	다운로드

							분석결	<mark>코</mark> 라 (1-mode) - 유클리(디언계수					
					메르스	코로나	사	4	바이러스		것	1	9	щ	-
			_	멧르	0 0	.901941932431	0.79148	5585943 0	893399641	822 0.8143	30466182	3 0.78679	9283644	0.764297	73
			_			군식결과(1*	moae) - 🗅	고자인세수					65	0.6666666	56
				메르스	코로나	사스		바이러스		것		19	<u> </u>	0.732738	7:
			메르스	0	0.290554035331	0.4539671226	29 0.2	19381727238	0.253	073177335	0.2078	319541703	64	0.7	ć
					분석결과 (1-mode	e) - 자카드계수						80261368	13	0.6666666	56
												19937932	-18	0.622035	52
			메르스	코로나	사스	바이근	<u>~</u>	것		19	^	175671644	91	0	
	_	메르 스	0	0.205933682373	0.292279411765	0.1685393	25843	0.150588	35294	0.1152882	20551	82402405	13	0	-
				분석결과 (1-mode) -	· 상관계수					0.2263157	/89474	0		0.591751	-
										0.0940170	94017	61064764	83	0.292893	2'
	메르스		코로나	사스	바이러스		것		19	0.0409523	72424	343035813	5	0 591751	7(
멧르	0		0.206616444608	0.427450639988	0.1278515238	.061	20821247	1 0.046	178492	0.04989333	07243-	01421445		0	
코로	0.20661644	4600	0	0 101000142575	0 2202067600	0.0447	05700244	4 0.513	002016/	0.0732758	362069	974725666	6 5	0.0	,
나	0.200010444	4008	0	0.181898143575	0.3292967609	0.0447	05700341	4 0.512	0030104	0	24.704	98772628		0.0	,
사스	0.427450639	9988	0.181898143575	0	0.2633464451	0.152	3660795	7 0.023	2936769	0.0717948	/1/94	96463656	81	0.0	-
바이 러스	0.127851523	3891	0.329296760968	0.263346445198	0	0.0578	30776781	2 0.064	7122687	0.0326086	95652	763156939	_ +	0 202803	2.
것	0.061208212	2471	0.0447057003414	0.152636607957	0.0578307767	812	0	0.013	9968784	0.0402298	85057	292933435	0	0.0	-
19	0.046417849	92521	0.512883816455	0.0232936769613	0.0647122687	901 0.0139	96878473	32	0	0.0402250		344093752	(È
αH	0.151405477	7366	0.0368247070502	0.0987442733342	0.1199767862	0.0618	03507738	6 0.037	320778	0.1190476	519048	15381475	C		
개발	0.057790681	0547	0.0336550363362	0.108759776196	0.00774767687	0.102)2227136	5 0.040	109731	0.0783132	253012	396201286	C		
수	0.039221547	77494	0.0100786613419	0.0873963839952	0.00619315656	0.0273	45685165	0.113	0782283	0.0368098	15950	97263725	-		
틍	0.084949146	59634	0.109104771272	0.0463139354076	0.0534878608	712 0.0971	07022511	4 0.079	3596967	0.0184049	07975	-			
신종 플루	0.118867016	6678	0.143516107607	0.0437422418514	0.0732097652	381 0.0396	28654319	0.206	0261259	0.0122699	38650				
사태	0.130549358	8618	0.0715859381448	0.099353276482	0.1013678278	0.0235	86867966	0.039	5405557	0.0880503	14465				
신종	0.039600261	5727	0.253971550584	0.201218929616	0.3628049298	0.0830	94038518	0.060	7407984	0.0387096	77419				
백신	0.041066380)5253	0.0408732533669	0.0242277040026	0.0315193387	276 0.0218	12341537	0.024	7677223		+				
치료 제	0.023793775	53954	0.133141305234	0.060424094882	0.0449446861	563 0.0333	26511818	0.088	0147805						
2015 년	0.106687845	5169	0.161236803114	0.132107451181	0.0177024890	188 0.117	13699909	5 0.070	5332533	L					
감염	0.035664249	93662	0.0113356670685	0.0125840441496	0.1659059461	0.0485	17563913	0.036	1299124						
4															

매트릭스 만들기



학습데이터 만드는 법 자세히 알아보기

- ② 학습데이터 전체 데이터를 분류하기 위한 기준이 되는 데이터로, 각 극성의 비율을 비슷하게 태깅하여 업로드해줍니다.
- 원문데이터 감성분석의 분석결과를 확인하려면 원문데이터를 다운로드(Excel)하여 학습데이터를 만들어야 합니다

원문데이터에 직접 극성을 부여한 학습데이터를 업로드하여, **문서의 극성 분석결과**와 **감성단어 빈도 분석결과**를 확인할 수 있습니다

								•
техтом	≞ 감성	분석					<u>데이터 용량추가</u>	ଗ୍ରତାଖିଲ୍ୟା 🗸 👤
● 수집하기	메르스	Q 검색결과 23 / 21548			≡ 107╫ ✓ >	텍스트마이닝 매트릭스	감성분석	토픽분석
▲ 데이터전처리 ✓ 수집리스트	포털/SNS	뉴스 보유데이터 요청채널			데이터적용리스트	원문데이터 미리보기	다운로드(Excel)	다운로드(txt)
정제/형태소분석		데이터명		생성날짜	용량	2 학습데이터	학습	헤이터가 적용 되어있습니다.
☴ 데이터분석 🗸 🗸		코로나 +메르스	Ľ	2020-03-20	6.26 MB	🗴 예시파일 다운로드		
테스트마이니		코로나 +메르스 [수집단위] 2020-03-13 ~ 2020-03-13	ľ	2020-03-20	6.26 MB	예시파일을 참고하여 최소 10 으로 그녀하여 어르도 해주세?)건에서 최대 1,000건의	데이터를 긍정/중립/부정
		Bit to	8	2010/08	3019	(비율을 비슷하게 업로드 할 수	록 정확한 결과가 나옵니	.[다.]
		(b.n.	2	2010/00	1754	브성경과 /모 러기비아		
감정문식		Rear .		000000	1752	· 전체 (금정/중립/부정)		
토픽분석		425.028		1004010	2010	미리보기		
교 시각화 🗸		No. 108		2010/07	38,2654			-12
		Ref. n.	8	2014-014	1.0	· 중성		
시각화결과		- the second sec	2	2010-0-28	274	비리보기		주기분석
커스터마이징		- Alexandrian Contraction (Contraction)	10	2010/11/20	5.0000	› 중립		
						미리보기	다운로드	추가분석
) 부정		
						미리보기		추가분석
						• 감성단어 빈도		
						미리보기		다운로드(txt)
						미리보기 일부 데이터를 확인	할 수 있습니다. <u>전체 데</u>	이터를 원하는 경우 다운로

감성분석 하기

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

어들에 대하여 개별적인 분석을 하고 싶은 경우, 추가분석 기능으로 가능합니다

36.06 40.54

분석결과 학습데이터가 적용되면 분석결과를 확인할 수 있습니다 ● 문서의 극성 분석 학습데이터 기준으로 분류된 <u>전체 문서의 극성</u>을 확인할 수 있으며, 긍정·부정·중립으로 분류된 단

감성분석 결과 (문서기반) - 전체									
구분	빈도(건)	1	비율(%)						
전체									
긍정	225								
중립	253								
부정	146								
	분석문장		결과						
코로나(covid-19) 검사 후기 코로 럼 나	나검사에 더 겁먹은 왕쫄보라 덜덜 떨며 선별진료소로 향했다.	코로나는 사스,메르스처	중립						
특징 기사 1코로나 19의 현황과 전망 중중까지 다양하고 사망에 이를 수도 있다고 전했다. 다른 자료에 따르면코로나 19는 매르스또는 사스에 비교해중									
세계가 주목하는코로나19 검체법 르스때에 이어	t, 워킹스루 진료소 귀감이 되었어요. 드라이브스루보다 더 건	'편해진 워킹스루 도입!메	긍정						
알고 보면 절대 무섭지 않은 바이 대의 높은 치사율	러스의 진실 / 이재갑 교수 (바이러스 경우 비교적 중환자 치료	리가 잘 되었음에도 20%	중립						
코로나19 자발적 자가격리 한 달 모든 국경을 차단시키	째~육아도 이제 일상이다. 활동한지 벌써 15년이 넘었는데 이빈	번코로나19처럼 전 세계가	부정						
코로나가 진짜 무서운 이유 사스며 우 심합니다. 쭉쭉 오르네요	베르스신종플루 에볼라코로나19 그래프와 같이코로나가 진짜 *	무서운 이유는 전염성이 매	긍정						
[공지]코로나19 종식 희망 릴레이 희망	-코로나19로 인해 얻은 것들 님께서 감사하게 blog.naver.cc	om코로나19 종식을 위한	긍정						
의성마늘 역고코로나퇴치하자 생마늘 고깃도 의성마늘 역고코로나초천박살 따봉입니다 . 마늘은 한국인의 대표 넣으니 정답이고 말고다									
2020년 방송통신대 1학기 중간과제을 경리 방송통신] 9 page 2020년 1학기 보건의사소통 중간시험과제물 C형(코로나 19와메르스확									

- C	30042-00			10	10 A		
10	1000000				10.00	분석결과 (문서기반)	
	101-01-04			304		▶ 전체 (긍정/중립/부정)	
						미리보기	다운로드(Excel)
	1000					, 구제	
	형태소 분석 결과 대	미 <mark>리보기</mark> (분석품사)				0.0	
					- C	미리보기	다운로드
구분	빈도(건)	감성강도비율(%)	ы Ы	도비율(%)	10) 중립	
긍정	327 / 588	54,27 /	100.0	55,61 / 100.0		미리보기	다운로드
부정	261 / 588	45,737	100.0	44,39 / 100.0		, 비거	
						7 10	
키워드 부정키워드 세부	감성					미리보기	다운로드
키워드 부정키워드 세부 	감성					미리보기	다운로드
키워드 부정키워드 세부감 - 호감 기쁨 감정분류	감성 빈도(ඊ)	감정강도	빈도 * 감성강도	빈도비율(%)		미리보기 · 감성단어 빈도 미리보기	다운로드
키워드 부정 키워드 세부 - 호감 - 기쁨 - 김경분류 해다	감성 빈도(2) 11	감청강도 3.77778	빈도 * 감성강도 41.55558	변도비율(%) 1.87		· 구공 미리보기 · 감성단어 빈도 미리보기	다운로드
키워드 부정 키워드 세부, 호감 기름 김정분류 하다	감성 빈도(2) 11 8	감정강도 3.77778 2.7778	변도 • 감성강도 41.55558 22.2224	빈도비율(%) 1.87 1.36		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 확 드 기능을 이용하세요.	다운로드 밖인 할 수 있습니다. 전체
부정키워드 세부: 호감 1 기름 합하다	방성 번도(2) 11 8 6	<mark>감청강도</mark> 3.77778 2.7778 4.66667	변도 * 강성강도 41.55558 22.2224 28.00002	벤도비율(%) 1.87 1.36 1.02		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 해 드 기능을 이용하세요. 추가분석 형태소 분석, 네!	다운로드 바인 할 수 있습니다. 전체 트워크 매트릭스 작성 등
기위드 부장기위드 세부: - 호강 기름 	방성 번도(2) 11 8 6 3	감정감도 3.77778 2.7778 4.66667 3.88889	변도*강성강도 41.55558 22.2224 28.00002 11.66667	빈도비율(%) 1.87 1.36 1.02 0.51		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 확 드 가능을 이용하세요. 추가분석 형태소 분석, 네 습니다. 감성분석 시각화 결파는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등· I2화결과 페이지에서 확
기원도 부정기원도 세부: 호강 기봉 강경분류 하다 하다 기적 격이다	방성 번도(2) 11 8 6 3 3	28825 3.7778 2.7778 4.66667 3.88899 3.3333	번도 * 감성경도 41.5558 22.2224 28.00002 11.6667 6.6666	벤도비율(%) 1.87 1.36 1.02 0.51 0.34		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 第 드 가능을 이용하세요. 추가분석 형태소. 본석, 네 습니다. 감성분석 시각화 결과는 시	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등 각화결과 페이지에서 확
기역도 부정기역도 세부: 호감 기봉	방성 변도(2) 11 8 6 3 3 2 2 1	김정갑도 3.77778 2.7778 4.66667 3.88889 3.3333 5.0	번도 * 감성감도 4155588 222224 28.0002 11.66667 6.6666 5	번도비율(%) 1.87 1.36 1.02 0.51 0.34 0.17		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 확 드 기능을 이용하세요. 추가분석 형태쇼 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등· 각화결과 페이지에서 확
기원도 부장 키워드 세부 호값 기름 건경분류 - - 감정 - - 태다 - - 13 - - 14 - - 15 - - 14 - - 15 - - 14 - - 15 - - 14 - - 15 - - 15 - - 16 - - 17 - - 18 - - 19 - - 19 - - 19 - - - 19 - - - 19 - - - 19 - - - 19 - - -	방성 변도(2) 11 8 6 3 3 2 2 1 1 1	2/87/25 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778	번도 * 감성감도 41.55588 22.2224 28.0002 11.66667 6.6666 5 5 5.77778	번도비율(%) 1.87 1.36 1.02 0.51 0.34 0.17 0.17		· 구경 미리보기 · 감성단어 빈도 미리보기 미리보기 일부 데이터를 확 드 기능을 이용하세요. 추가분석 형태쇼 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 착성 등 각화결과 페이지에서 확
기원도 부장 키워드 세부 호감 기응 감정분류 - - 감정 - -	방성 변도(전) 111 8 8 6 3 3 2 2 1 1 1 1 1	2/8/2/5. 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0	번도 * 감성감도 41.55588 22.2224 28.0002 11.66667 6.6666 5 5.77778 4	번도비율(%) 1.87 1.36 1.02 0.51 0.34 0.17 0.17 0.17		- 구경 미리보기 - 김성단어 빈도 미리보기 일부 데이터를 된 드 기능을 이용하세요. 추가분석 힘태소 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등 각화결과 페이지에서 확
기원도 부장 키원도 세부 호감 기응 감정분류 - - 감정분류 - - 감정 - - - - </td <td>방성 변도(전) 111 (111) (11)</td> <td>2/87/25 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0 4.6667</td> <td>번도 * 감성감도 41.55588 22.2224 28.00002 11.66667 6.6666 5 5.577778 4 0</td> <td>변도비율(%) 1.87 1.36 0.51 0.34 0.17 0.17 0.17 0.17</td> <td></td> <td>· 구경 미리보기 · 감성단어 빈도 미리보기 일부 데이터를 될 드 가능을 이용하세요. 추가분석 힘태소 분석, 네 습니다. 감성분석 시각화 결과는 사</td> <td>다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 착성 등 각화결과 페이지에서 확</td>	방성 변도(전) 111 (111) (11)	2/87/25 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0 4.6667	번도 * 감성감도 41.55588 22.2224 28.00002 11.66667 6.6666 5 5.577778 4 0	변도비율(%) 1.87 1.36 0.51 0.34 0.17 0.17 0.17 0.17		· 구경 미리보기 · 감성단어 빈도 미리보기 일부 데이터를 될 드 가능을 이용하세요. 추가분석 힘태소 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 착성 등 각화결과 페이지에서 확
기원도 부장 키원도 세부 호감 기응 감정분류 - - 감정분류 - - 감정 - - - - </td <td>방성 변도(전) 원동(전) 111 8 8 8 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7</td> <td>2/87/25 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0 4.6667 3.3333</td> <td>번도 * 감성감도 41.5558 22.2224 28.0002 11.6667 6.6666 5 5.77778 4 0 0</td> <td>변도비율(%) * * * * * * * * * * * * * * * * * * *</td> <td></td> <td>· 구경 미리보기 · 감성단어 빈도 미리보기 일부 데이터를 될 드 가능을 이용하세요. 추가분석 힘태소 분석, 네 습니다. 감성분석 시각화 결과는 사</td> <td>다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등· 각화결과 페이지에서 확</td>	방성 변도(전) 원동(전) 111 8 8 8 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	2/87/25 3.77778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0 4.6667 3.3333	번도 * 감성감도 41.5558 22.2224 28.0002 11.6667 6.6666 5 5.77778 4 0 0	변도비율(%) * * * * * * * * * * * * * * * * * * *		· 구경 미리보기 · 감성단어 빈도 미리보기 일부 데이터를 될 드 가능을 이용하세요. 추가분석 힘태소 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드 안인 할 수 있습니다. 전체 트워크 매트릭스 작성 등· 각화결과 페이지에서 확
기원도 부장 기원도 세부 호감 기품	방성 변도(2) (11) (33) (33) (33) (33) (33) (33) (33	2/87/25. 2.7778 2.7778 4.66667 3.88889 3.3333 5.0 5.77778 4.0 4.6667 3.3333 3.333 2.4444	번도 * 감성감도 41.55588 22.2224 28.00002 11.66667 6.6666 5.577778 4 0.0 0 0 0	변도비율(%) 1.87 1.36 0.51 0.34 0.17 0.17 0.17 0.17 0.17 0.17 0.17 0.17		· 구경 미리보기 · 감성단어 빈도 미리보기 일부 데이터를 해 드 기능을 이용하세요. 추가분석 할택소 분석, 네 습니다. 감성분석 시각화 결과는 사	다운로드

감성분석 하기

④ 분석결과 적용이 완료되면 연관성이 높은 단어들끼리 군집으로 분류가 된 결과를 확인할 수 있습니다

- ❸ **적용** 군집 수(K값)과 군집별 단어 수를 설정 후 적용 버튼을 클릭합니다
- ⑦ 군집별 단어 수 군집 내에 들어갈 단어의 개수를 적어줍니다
- 군집 수(K값) 만들고 싶은 군집의 개수를 적어줍니다

토픽분석 하기

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

техтом	글 토픽분석						<u>데이터 용량추가</u> 더아이엠씨님 ~ 👤
			분석결과 (텍스트마이닝)				
을 수집하기	코로나 +메르스	데이터명	생성날짜	용량	• •	텍스트마이닝 매트릭스	감성분석 토픽분석
	비역세	코로나 +메르스	2020-05-12	6.26 MB		Word-level Semantic	Clustering
▲ 네이터전서리	포털/SNS 뉴스	상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운	로드하여 확인할 수 있습니다.	다운로드		문서 내 단어를 임베딩하여 단어 Hierarchical Clustering을 통하	벡터 간의 묶음 관계를 Agglomerative 확인하는 방법입니다.
구입니스트		그룹	단어	Topic Modeling		군집 수가 사용자가 지정한 것 보	다 작을 경우 임의 지정 됩니다.
정제/형태소분석		1	코로나	277		▶ 군집 수 (K 값)	10 개
≔ 데이터분석 ✓		1	÷	73	бМВ	• 군집별 단어 수	20 71
		1	5	71	5 MB		2 ✓ 적용
텍스트마이닝		1	발생	37			
매트릭스		1	2015년	27			
감성분석		1	당시	23	1 1		
토픽분석		1	관련	23		문석결과	
뇨 시각화 ◇		1	전염병	23		Word-level Semantic Clu	우로드(Excel) 다우로드(tyt)
		1	사람	23			
시각화결과		1	3월	20		 LDA Topic Modeling 	
커스터마이징		1	오늘	15		미리보기 다	운로드(Excel) 다운로드(txt)
		1	위협	9			
		-		~ •			용 에부를 선택합니다. 용 네 트피 모델리 경기이 개왕서이 떨어진

3. 데이터뷴

당 분석결과 적용이 완료되면 키워드가 어떤 그룹으로 분류가 되었는지 확인할 수 있습니다

④ 적용 토픽 수와 토픽 단어 수, 랜덤 값 설정 후 적용 버튼을 클릭합니다.

니다. 분석 결과의 재현성을 확보하고 싶은 경우 사용안함을 선택해주세요

8 랜덤 값 무작위 할당을 진행할 경우 분석 결과의 재현성이 떨어지기 때문에, 같은 데이터로 분석을 진행하더라도 결과값이 달라질 수 있습

② 토픽 단어 수 그룹 내에 들어갈 단어의 개수를 적어줍니다

● **토픽 수** 만들고 싶은 그룹의 개수를 적어줍니다

교 시각

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

상상분석						
프로		문석결과 (텍스트마이닝)		LDA TOPIC MODE 토픽모델링은 대량의 문 악고리즈으로 유사하여	eling -서군으로부터 주제(토픽)을 이미를 가지 다어들을 진다히	· 자동으로 찾아내기 위한 하하는 방식으로 토필을 추
l각화 ~	데이터명	생성날짜	88	로 다음		10 7
각화결과	코로나 +메르스 상위 200개까지 타어록 미리 볼 수 있습	: 2020-05-12 	6.26 MB	2 → 토픽 단어 수		20 71
12clnto18	그룹	단어	Topic Modeling	3 • 랜덤 값	사용 샘플링 과정에 포함	사용안함 된 무작위 토픽 함당 기능
	1		0.024		사용 여부를 선택합! 사용 시 토픽 모델링 스 이스니다	-I다. 결과의 재현성이 떨어질
	1	사스	0.022		ㅠ ᆻ급니다. 학습데이터가 적용 되이	r <u>Na</u> ✓ 적용
	1	것	0.019			
	1	바이러스	0.015	티서겨고니		
	1	년	0.011	문격열파 ▶ Word-level Sema	ntic Clustering	
	1	19	0.009	미리보기	다운로드(Excel)	다운로드(txt)
	1	신종	0.008	► LDA Topic Model	ing	
	1	2015	0.007		다우르드(Excel)	

※ 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

- ⑦ 키워드 선정 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다
- 색상선택 4가지의 기본 색상 조합 또는 우측의 상, 중, 하에서 직접 원하는 색상을 선택할 수 있습니다

※ 상위 빈도 순으로 최대한 유의미한 키워드들을 선택하는 것이 좋습니다

워드클라우드는 문서의 키워드를 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 시각화로<mark>,</mark> 단어빈도 결과값을 사용합니다

техтом	는 시각화결:	과								<u> 터 용량추가</u> 더(아이엠씨님 🗸 🙁
🛢 수집하기	워드클라우드 ㅂ	바차트 에고네트워크	네트워크 개체명인식(트리민	법) 토픽모델링 클	러스터링 매트릭스 차트	문서 감성분석	키워드 감성분석	어휘 감성분석			
▲ 데이터전처리 🛛 🗸		분석단어		수집	입날짜	88	÷	_ 4	▲ 다운	로드	
		코로나 +메르	<u>.</u>	2020	-03-20	6.26	KB	클라우드 모양 변경	ł		
수집리스트								틀이 되는 이미지를	업로드하여 특정 모양!	으로 시각화를 생성할	발 수 있습니다.
정제/형태소분석								· 점을이미지을 잠고이 며 새푸이미지1 새	·여 원하는 모양의 이미 프이미지?	[시(Jpg, png)들 입의 	도드애 모세요.
									글이미지스		√ 적용
≔ 데이니군식											
텍스트마이닝								색상선택			
매트릭스				지어버							중 (하)
감성분석		일양의	분품 미국 대관	신엄명				2 키워드 선정			
토픽분석			약산 전문가 ^{사람}	인국	_			빈도	키워드	내림차순	오름차순
			^{중상} 사태 ^비	<u>,</u> ,,신종플	루			✔ 상위 30개	선택해제	숫자	영어
교 시각화 🔹			세계 유행	끡신 발생				1~5	0	51~100	0
시간화격과			개발		슈펙트			메르스	296 🔽	증상	23
								코로나	277	현재	23
지스티마이징		HE	ヘレトイン		효과			✓ 사스	248	후보물실 마	23
				중동호읍	기승우군			·····································	129	 슈펙트	22
			- 기-에버 과7	<mark> 치료제</mark> 전프	· 대응			년	110	일	22
									· · _ · 🛥	-01-21-71	• 22

시각화 결과 확인하기

※ 막대 위에 마우스를 올리시면 데이터 값을 확인할 수 있습니다

※ 상위 빈도 순으로 최대한 유의미한 키워드들을 선택하는 것이 좋습니다

여부를 선택할 수 있습니다 ❷ 키워드 선정 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다

■ 시크는 문지의 한이란도를 한도에 비데이는 물이의 국내로 다니면 지르답더니
 ● 시각화 설정 빈도는 막대의 색상을, 백분율은 점의 색상을 선택할 수 있으며, 데이터 표시는 차트 위 빈도, 백분율 값 표시



시각화 결과 확인하기

46



시각화 결과 확인하기

바차트

에고네트워크

네트워크

개체명인식(트리맵)

토픽모델링

클러스터링

매트릭스 차트

문서 감성분석

키워드 감성분석

어휘 감성분석

시각화결과

E

워드클라우드

TEXTOM

🛢 수집하기

<u>데이터 용량추가</u> 더아이엠씨님 ~ 📿

※ 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

- ⑦ 키워드 선정 빈도 또는 키워드(가나다) 기준으로 정렬할 수 있으며 중요하다고 생각되는 키워드들을 선택해줍니다
- 시각화 설정 선 색상은 화살표의 색상을, 화살표 모양은 화살표의 머리 모양을 선택할 수 있습니다

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



시각화 결과 확인하기

※ 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

● 개체명 사람, 학문, 대상물, 기관, 지역, 문명, 날<mark>짜, 시간 숫자, 사건/사고, 동물, 식물,</mark> 금속, 용어 <u>개체명 인식 자세히 알아보기</u>

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



시각화 결과 확인하기

49

- ③ 글자크기 설정 글자의 크기를 설정할 수 있습니다
- ❷ 시각화 설정 선의 모양을 설정할 수 있습니다
- 색상 설정 선 색상, 원의 색상, 글자의 색상을 설정할 수 있습니다

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



시각화 결과 확인하기

시각화 결과 확인하기



클러스터링은 Word-level Semantic Clustering 분석을 트리맵으로 나타낸 시각화입니다

(Word-level Semantic Clustering 분석을 진행해야 시각화 결과를 확인할 수 있습니다)

※ 전체토픽 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

4. 시각화 🖕



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

빨간 막대그래프는 해당 토픽에서의 빈도를 보여줍니다

⑤ 토픽 구성 단어 토픽을 구성하는 단어들을 확인할 수 있으며, 파란 막대그래프는 전체 단어의 빈도를,

④ λ(**람다) 값 설정** λ(람다) 값을 조절하여 토픽을 구성하는 단어의 출현 조건을 설정할 수 있습니다

가장 큰 원이 메인 토픽이라고 해석할 수 있습니다

❸ 토픽의 크기 토픽 원의 크기가 클 수록 높은 빈도수의 단어들로 구성되어 있습니다

토픽 간의 거리가 가깝거나 겹쳐져 있다면 판별 타당도가 낮음으로 비슷한 주제를 나타냅니다

② 토픽 간의 거리 토픽 간의 거리가 멀 수록 판별 타당도가 높고 주제가 뚜렷하게 구분됩니다

토픽을 구성하는 30개(사용자 설정 개수) 단어를 확인할 수 있습니다

LDA Topic Modeling 시각화 해석 ● 토픽선택 토픽분포도에서 토픽을 클릭하거나 토픽번호를 'Selected Topic'에 직접 입력하여 토픽을 선택하면 4. 시각회

※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

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

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



54

※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

※ 차트 위에 마우스를 올리시면 감성강도의 백분율 값을 확인할 수 있습니다

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



※ 학습데이터를 업로드하지 않으면 시각화가 나타나지 않습니다

※ 차트 계열 위에 마우스를 올리시면 세부감성비율 값을 확인할 수 있습니다

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



② 결과보기 업로드한 파일의 시각화 결과물을 확인할 수 있습니다

를 엑셀 파일로 업로드해줍니다....

시각화 커스터마이징하기

● 파일업로드 좌측 네모 박스 안의 정보와 우측 네모 박스 안의 Sample 데이터를 참고하여 각 시각화에 맞게 적합한 데이터

시각화 커스터마이징은 <u>무료</u>이며, 9개의 각 시각화별 양식에 맞는 파일을 업로드만 하면 원하는 시각화 결과물을 바로 확인할 수 있는 기능입니다

ТЕХТОМ	<u></u> 커스터마이징						데이터용량추가 더아이엠씨님 ~ 😩
을 수집하기	, ·		^{커스터마이징의 시각화 중} 보유하고 있는 데이터로	물과는 별도의 용량 소모 없이 무료 . 손쉽고 다양한 시각화	로로 제공됩니다. 를 완성해보세요.	0	
▲ 데이터전처리 🗸							
수집리스트	워드클라우드 바차트 에고!	네트워크 파이차트 라인차트 1	l-gram 네트워크 1-way 워드트리	트리맵 2-way 워드트리			
정제/형태소분석					0	_	
≔ 데이터분석 ✓	~ 말이닝테스톰	코로나+메르스_빈도100.xlsx			Ĭ	5	
텍스트마이닝	· · · · · · · · · · · · · · · · · · ·	아래 이미지와 같은 형식으로 된 빈도 엑셀	파일을 업로드 하시면 최대 100개 단어를 반영	하는 워드 클라우드가 생성됩니다		-	
매트릭스	제 성문소리 이야임자 전유 성관 사망 제 제 제가 다아이임자 활용 보고서 조로 이야 지기 예쁜 학습 문어 활용 보고서						
감성분석	fin 40	A	В	C D	E F		
토픽분석	워드클라우드	1 빅데이터	29871				
	A열 : 키워드(100개) R연 : 비도	2 분석	6118				
🖬 শব্বহ্য 🔹 👻		3 활용	3456				
시각화결과	2 결과보기	4 정보	2647			_	
커스터마이징		5 전문가	2408			_	
		<u>6</u> 기술	2086				
		7 산업	1911				
		8 기반	1576				
		9 연구	1544				
		10 교육	1431				