

통일 빅데이터 센터 자료팀 발표

박유경

자료분석 메뉴

- 매뉴얼
- 데이터 전처리
- 데이터 분석

자료 분석

메뉴얼

데이터 전처리

데이터 분석

데이터 전처리

- 내 보관함: 검색한 키워드에 해당하는 문서들을 선택하여 분석을 진행할 수 있다.
- 사용자 사전: 불용어, 유의어, 복합어
- 품사 선택: 명사, 형용사, 동사 중 선택하여 전처리를 진행할 수 있다.

데이터 분석 및 시각화

대상 문서에서 추출한 top N개의 주요 키워드들을 볼 수

-빈도수분석
(설명 팝업)

-TFIDF

-N-gram

-유의어분석

-군집분석

-의미연결망

-토픽모델링

-개체명 인식

- 표, 워드클라우드, 바차트
- 표, 워드클라우드, 바차트
- 목록, 네트워크 그래프
- 표, 동시출현맵
- 목록, 산포도
- 표, 네트워크 그래프
- 목록, 토픽 그래프
- 목록

상세매뉴얼 다운로드

데이터 전처리

자료 분석

메뉴얼

데이터 전처리

데이터 분석

[내 보관함]

검색 키워드	문서목록	수집날짜	통계	전처리
<input checked="" type="checkbox"/> 북한 + 통일	북한이탈주민의 통일에 대한 인식 조사	2021-3-30	총 문서수 : 3개 평균 단어 수 : 230개	
	통일을 위한 초등교육			
	통일을 위한 북한의 언어와 남한의 언어의 차이 연구			
<input type="checkbox"/> IT	북한의 IT기술	2021-2-28	총 문서 수: 1개 평균 단어 수: 100개	

사용자 사전

유의어 ☒ 사용함 ☐ 사용안함

불용어 ☒ 사용함 ☐ 사용안함

복합어 ☒ 사용함 ☐ 사용안함

사용자사전 파일명

[미리보기](#)

품사 선택 ☐ 동사 ☒ 명사 ☐ 형용사

[전처리 실행](#)

데이터 전처리

자료 분석

메뉴얼

데이터 전처리

데이터 분석

[내 보관함]

검색 키워드	문서목록	수집날짜	통계	전처리
<input checked="" type="checkbox"/> 북한 + 통일	북한이탈주민의 통일에 대한 인식 조사	2021-3-30	총 문서수 : 3개 평균 단어 수 : 230개	<div>미리보기</div> <div>다운로드</div>
	통일을 위한 초등교육			
	통일을 위한 북한의 언어와 남한의 언어의 차이 연구			
<input type="checkbox"/> IT	북한의 IT기술	202		

전처리 실행버튼을 누르면
전처리 결과를 미리보거나 csv파일을
다운로드 받을 수 있다.
전처리 결과는 각 문서별 추출된
키워드 리스트이다.

사용자 사전

유의어 ☒ 사용함 ☐ 사용안함

불용어 ☒ 사용함 ☐ 사용안함

복합어 ☒ 사용함 ☐ 사용안함

사용자사전 파일명

품사 선택 ☐ 동사 ☒ 명사 ☐ 형용사

데이터 분석

자료 분석

[내 보관함]



검색 키워드	문서목록	수집날짜	통계	전처리
북한 + 통일	북한이탈주민의 통일에 대한 인식 조사	2021-3-30	총 문서수 : 3개 평균 단어 수 : 230개	미리보기 다운로드
	통일을 위한 초등교육			
	통일을 위한 북한의 언어와 남한의 언어의 차이 연구			
IT		-28	총 문서 수: 1개 평균 단어 수: 100개	

▶ 클릭하면 아래로 드롭다운 화면이 펼쳐져서 옵션사용하기 할 수 있음.
(??) 클릭하면 매뉴얼 팝업창

▶ 빈도수분석 (??)

▶ N-gram (??)

▶ TFIDF (??)

▶ 유의어분석 (??)

▶ 군집분석 (??)

▶ 의미연결망 (??)

▶ 토픽모델링 (??)

▶ 개체명인식 (??)

메뉴얼

데이터 전처리

데이터 분석

전처리를
실행하지 않은
문서는 접근할 수
없도록 구현

자료 분석

메뉴얼

데이터 전처리

데이터 분석

[내 보관함]



검색 키워드	문서목록	수집날짜	통계	전처리
북한 + 통일	북한이탈주민의 통일에 대한 인식 조사	2021-3-30	총 문서수 : 3개 평균 단어 수 : 230개	<div>미리보기</div> <div>다운로드</div>
	통일을 위한 초등교육			
	통일을 위한 북한의 언어와 남한의 언어의 차이 연구			
IT	북한의 IT기술	28	총 문서 수: 1개 평균 단어 수: 100개	

▶ 클릭했을 때 펼쳐지는 드롭다운 화면

▼ 빈도수분석 (??)

상위 100 개

실행

▶ N-gram (??)

▶ TFIDF (??)

▶ 유의어분석 (??)

▶ 군집분석 (??)

▶ 의미연결망 (??)

▶ 토픽모델링 (??)

▶ 개체명인식 (??)

데이터 분석

자료 분석

메뉴얼

데이터 전처리

데이터 분석

[내 보관함]

검색 키워드	문서목록	수집날짜	통계	전처리
북한 + 통일	북한이탈주민의 통일에 대한 인식 조사	2021-3-30	총 문서수 : 3개 평균 단어 수 : 230개	미리보기 다운로드
	통일을 위한 초등교육			
	통일을 위한 북한의 언어와 남한의 언어의 차이 연구			
IT	북한의 IT기술	2021-2-28	총 문서 수: 1개 평균 단어 수: 100개	

▶ 빈도수분석 (??)

▶ N-gram (??)

▶ TFIDF (??)

▶ 유의어분석 (??)

▶ 군집분석 (??)

▶ 의미연결망 (??)

▶ 토픽모델링 (??)

▶ 개체명인식 (??)

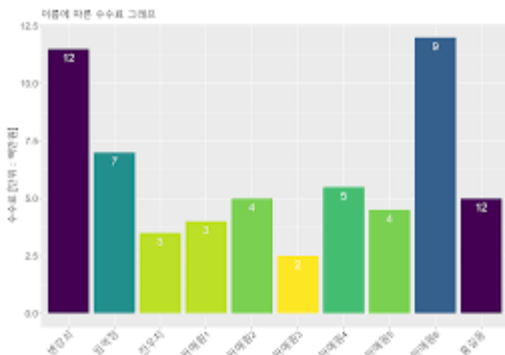
자료분석 시간이 많이 걸려서 사용자가 기다리지 않고 페이지를 떠났을 경우를 대비하여, flask에서 분석이 끝나면 분석함 자동저장 할 것

앞 페이지에서 실행을 누르면 결과 csv파일 또는 그래프 이미지파일 생성

화면 아래에서 바로 미리보기 내분석함 저장, 다운로드 가능

키워드	빈도수
INFORMATION	12
RESEARCH	7
LARGE	3
STORAGE	3
SETS	4
STORAGE	5
SETS	4
SETS	9
SETS	12

다운로드



다운로드



다운로드



Thank
you