

텍스트 네트워크 분석 서비스

CYRAM Text Network Analysis Service

왜, 텍스트 네트워크 분석이 필요할까요?

수 천개의 논문 데이터에서 연구 동향을 빠르게 파악하기 위해
특정 이슈와 관련해 트위터, 뉴스기사, 블로그에서 여론을 파악하기 위해
우리조직에 쌓여있는 수많은 보고서를 가치있게 활용하기 위해

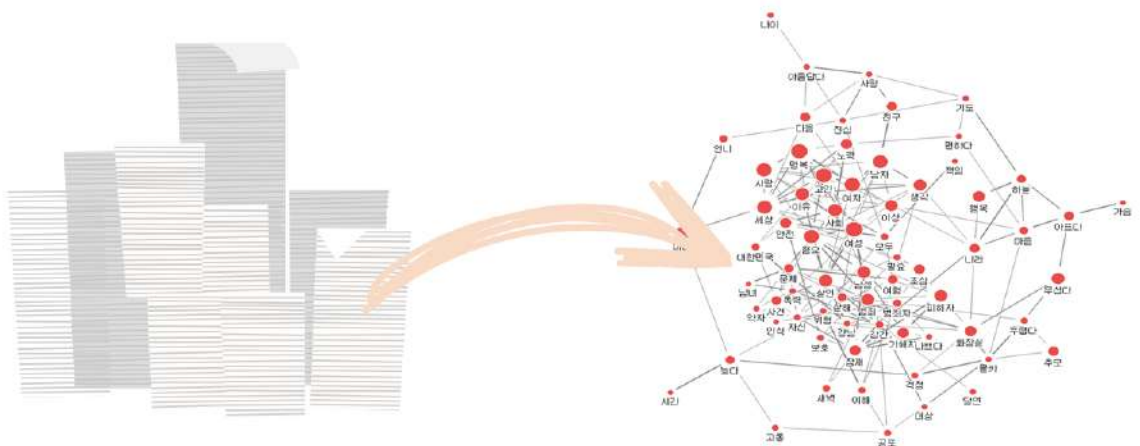
하지만, 텍스트 네트워크 분석은 어렵습니다.

키워드 추출은 어떻게 하지? 많이 등장하는 키워드가 좋은건가?
키워드 지도를 어떻게 시각화 하는 것이 효과적이지?
키워드 지도는 어떻게 해석하지?

사이람 텍스트 네트워크 분석 서비스는 이런 분들에게 솔루션을 제공합니다!

사이람 텍스트 네트워크 분석 서비스는 논문, 신문기사, 문서 등의 데이터에서 소셜 네트워크 분석 기법을 통해 핵심 키워드를 추출하고 세부 토픽을 분석하여 제공하는 서비스입니다.

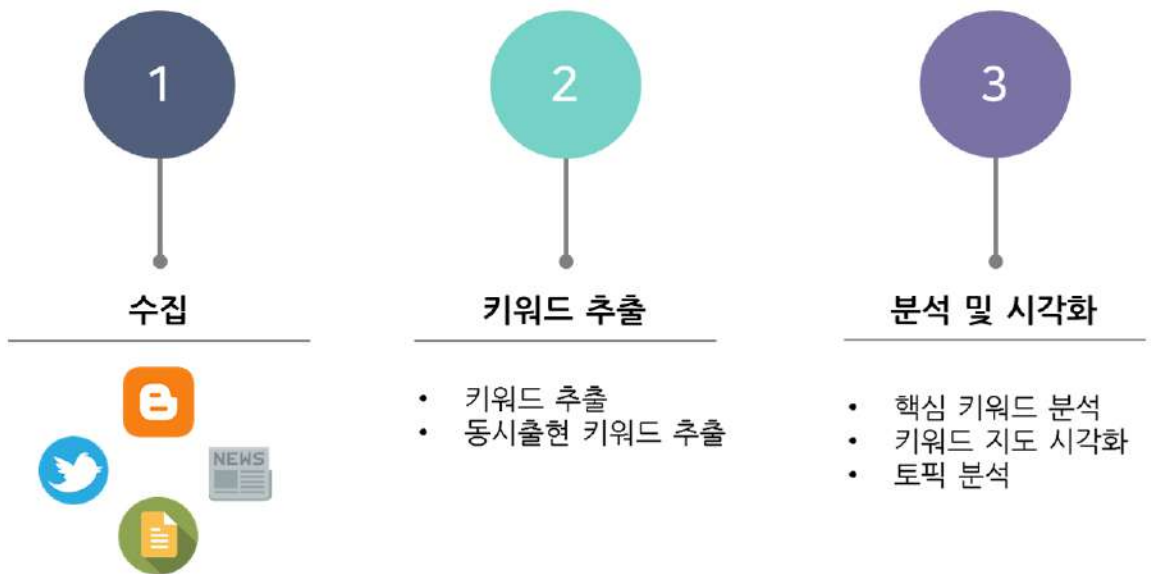
복잡한 데이터, 많은 데이터를 단순하고 직관적인 데이터로 재구성하여 연구동향 파악, 여론 분석에 효과적으로 활용할 수 있습니다.



Big Data, Complex Data

Simple & Visual

<텍스트 네트워크 서비스 개요 >



1. 데이터 수집

- 데이터 수집이 필요한 경우, 트위터/ 뉴스기사/ 블로그에 한해 수집 서비스를 제공해 드리고 있습니다.
- 그 외 데이터의 경우 협의가 필요합니다.

2. 키워드 추출

- 비정형 텍스트 문서에서 형태소 분석을 통해 키워드를 추출합니다.
- 필요에 따라 명사, 형용사, 동사, 부사 등을 선택적으로 추출할 수 있습니다.

<영문 키워드 추출 예시>

키워드	품사	등장빈도
America	Proper Noun	18
country	Common Noun	12
American	Adjective	11
have	Verb	11
nation	Common Noun	11
not	Adverb	10
again	Adverb	9

<국문 키워드 추출 예시>

키워드	품사	등장빈도
국민	Common Noun	57
시대	Common Noun	23
행복	Common Noun	20
경제	Common Noun	19
문화	Common Noun	16
새롭다	Adjective	16
국가	Common Noun	12

3. 동시출현 키워드 추출

- 함께 등장한 키워드 관계를 추출합니다.
- 추출단위는 문장, 문단, 문서 등 고객의 필요에 따라 선택할 수 있습니다.
- 필요에 따라 동시출현 빈도 또는 유사성 값을 모두 측정할 수 있습니다.

<문장 단위 동시출현 키워드 & 등장빈도>

키워드 1	키워드 2	등장빈도
국민	행복	25
국민	시대	21
시대	행복	20
경제	창조	13
국가	국민	10
가치	문화	9
국민	문화	9

<문장 단위 동시출현 키워드 & 유사성 값>

키워드 1	키워드 2	유사성 값
고난	역경	1
고난	현대사	1
고립	기대	1
과학	기술	0.8
교육	시스템	0.6667
발휘	역량	0.6667
역사	위대	0.6667

4. 핵심 키워드 분석

- 문서에 자주 등장한 핵심 키워드를 분석합니다.

- ① 빈도 기반 핵심 키워드: 여러 문서에 자주 등장한 키워드를 분석합니다.
- ② 중심성 기반 핵심 키워드: 중심성이 높은 키워드를 분석합니다. 중심성이 높을수록 주변 키워드를 서로 연결해주는 중심 키워드로서의 의미를 갖게 됩니다.

주제: 강남역 살인사건

<빈도 기반 핵심 키워드>

순위	키워드	출현 빈도
1	여성	293
2	명복	281
3	고인	272
4	여자	228
5	혐오	135
6	사람	122
7	세상	119
8	남자	119
9	미안	111
10	사회	99

: 여성, 명복, 고인, 여자 등 주제와 관련한 일반적인 키워드가 핵심 키워드로 추출됨

<중심성 기반 핵심 키워드>

순위	키워드	중심성
1	범죄자	0.1251
2	모두	0.108
3	화장실	0.0879
4	나라	0.0846
5	강간	0.0845
6	살인	0.0803
7	세상	0.0757
8	사회	0.0683
9	마음	0.0603
10	진심	0.0597

: 범죄자, 모두, 화장실, 강간 등 두려움에 관한 키워드가 핵심 키워드로 추출됨

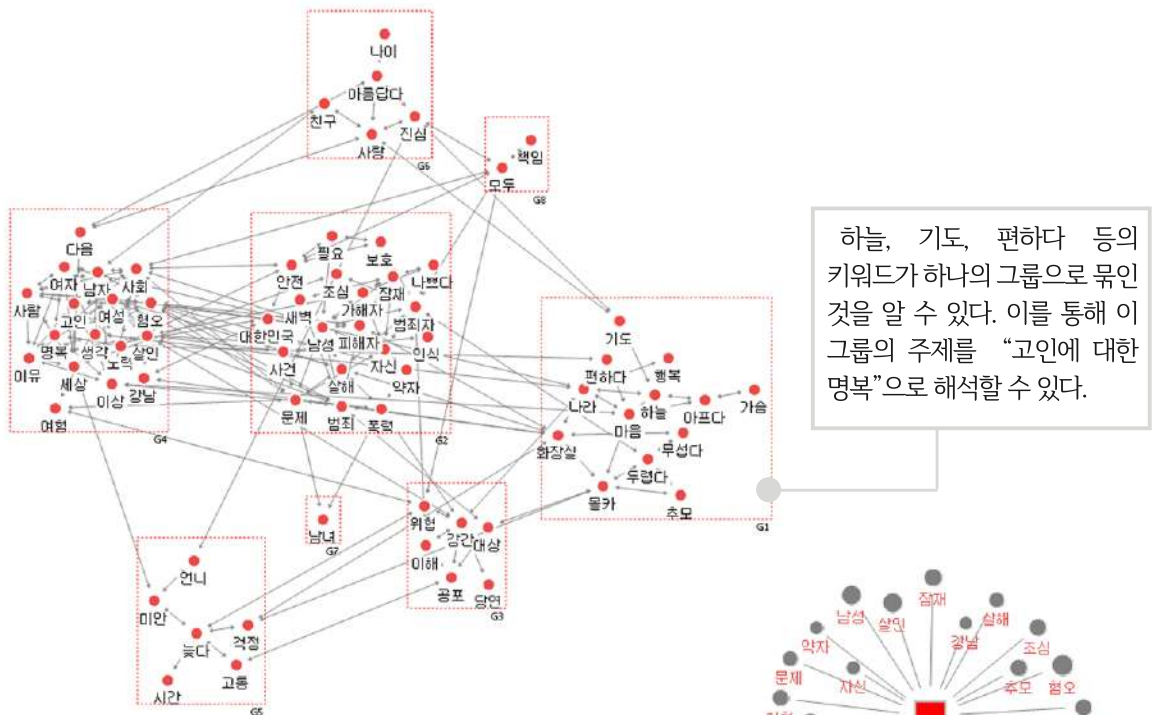
6. 토픽 분석

- 여러 문서에 내포되어 있는 공통 토픽을 분석합니다.

① 네트워크 기반 토픽 분석: 키워드 연결관계에 기반해 자주 함께 등장한 키워드 그룹을 토픽으로 추출합니다. 문서의 전체적인 개요를 파악하는데 효과적인 분석입니다.

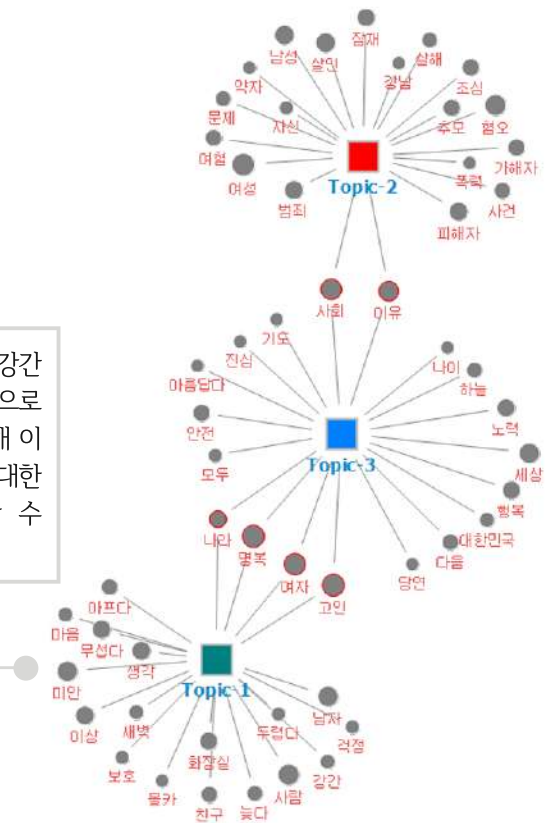
② 머신러닝 기반 토픽 분석: 키워드가 문서에 등장한 패턴을 분석하여 유사한 패턴을 가진 키워드들을 토픽으로 추출합니다. 문서별로 핵심 토픽을 파악하는데 효과적인 분석입니다.

<네트워크 기반 토픽 분석>



두렵다, 무섭다, 여자, 남자, 강간 등의 키워드가 하나의 그룹으로 묶인 것을 알 수 있다. 이를 통해 이 그룹의 주제를 “남성 범죄에 대한 여성들의 두려움”으로 해석할 수 있다.

<머신러닝 기반 토픽 분석>



FAQ

Q. PDF 파일도 분석이 가능한가요?

- CSV나 TXT 파일만 분석 가능합니다.

Q. 트위터/블로그/뉴스기사 모두 한꺼번에 분석할 수있나요?

- 동일한 주제이더라도 미디어에 따라 문서의 길이 및 내용 구성이 크게 다르기 때문에, 미디어 별로 각각 분석하는 것이 좋습니다.

Q. 데이터 용량이나 문서 양에 따라 서비스 가격이 달라지나요?

- 서비스 가격은 분석하고자 하는 주제와 미디어의 수에 따라 달라집니다.
자세한 비용은 이메일로 문의 주시기 바랍니다.

Q. 서비스 가격이 부담스럽습니다.

- NetMiner 텍스트 분석 패키지를 구입하시면 저렴한 비용으로 키워드 네트워크 분석을 수행하실 수 있습니다.

NetMiner 이용법 또는 텍스트 분석 방법이 낯선 분들을 위한 교육 서비스도 준비되어 있습니다.

CONTACT US

※ 아래 내용을 이메일로 보내주시기 바랍니다.

- ▶ 소속기관/이름/연락처
- ▶ 분석하고자 하는 주제
- ▶ 분석하고자 하는 데이터
- ▶ 필요 서비스
- ▶ 기타 남기고 싶은 말



E-mail consulting@cyram.com
Tel 031-739-8352