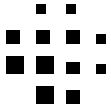


Topic Modeling
고독사

숭실대학교
산업공학과

AI &
HEALTHCARE LAB

01

AI &
HEALTHCARE LAB

데이터 수집

Crawling

- **연구 설계**
  - 연구 목표: 고독사 관련 뉴스 기사 수집 및 네트워크 분석(토픽 모델링 중심)
  - 검색 단어: 고독사, 고립사, 홀로 숨진 채 발견, 무연사, 무연고사망, 시간이 지나 발견된 시신
  - 검색 기간: 2010년 1월 1일 ~ 2023년 5월 31일
  - 검색 대상 신문사 (54개): 아시아경제, 아주 경제, 파이낸셜 뉴스, 한국경제, 헤럴드 경제, 강원도민일보, 강원일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주 매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부 매일, 중부일보, 충북일보, 충청일보, 충청 투데이, 한라일보, KBS, MBC, OBS, SBS, YTN, 디지털 타임스, 전자신문, 경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보, 매일경제, 머니 투데이, 서울경제
  - 전체 기사 수: 20742개

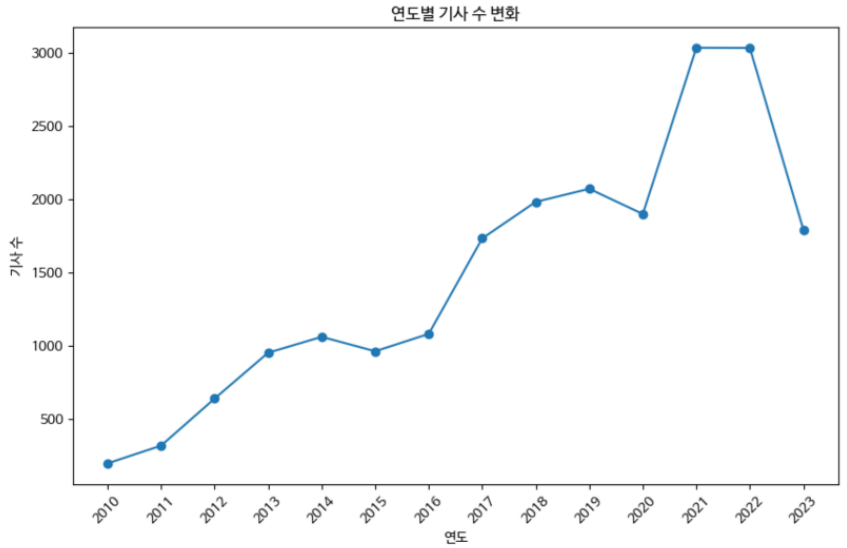


Fig. 연도별 기사 수 변화

데이터 처리

전처리

- **전처리 과정**
  - ① 기사 제목, 기사 본문을 데이터로 활용
  - ② 한글 추출
  - ③ 명사 추출
  - ④ 불용어 처리
  - ⑤ 길이가 2개 이상인 단어만 추출

01

AI &
HEALTHCARE LAB

분석

빈도 분석

- **상위 50개 키워드 빈도**
  - 총 단어 수: 3877355개

Rank	Text	Count	Rank	Text	Count	Rank	Text	Count
1	노인	45950	**18**	안전	12326	**35**	안부	9161
2	고독사	45821	**19**	추진	12038	**36**	상황	9012
3	사업	35219	**20**	확인	11956	**37**	진행	8533
4	가구	33570	**21**	관리	11369	**38**	취약	8046
5	복지	33187	**22**	건강	11007	**39**	죽음	8000
6	지원	26840	**23**	운영	10964	**40**	계층	7987
7	지역	26402	**24**	대상	10544	**41**	위기	7915
8	서비스	24165	**25**	활동	10296	**42**	증가	7888
9	예방	17756	**26**	이웃	10177	**43**	청년	7650
10	주민	17611	**27**	조사	9767	**44**	기자	7637
11	독거	17154	**28**	위험	9630	**45**	방문	7616
12	발견	17085	**29**	정책	9600	**46**	마을	7598
13	어르신	15632	**30**	계획	9591	**47**	사망	7501
14	생활	15045	**31**	경제	9377	**48**	발생	7477
15	센터	14722	**32**	제공	9327	**49**	고립	7460
16	가족	13988	**33**	발굴	9258	**50**	한국	7332
17	경찰	13177	**34**	서울	9222	-	-	-

01

AI &
HEALTHCARE LAB

분석

빈도 분석

- **빈도 워드 클라우드**
  - *고독사, 노인, 사업, 어르신, 가구*등의 단어는 전체 단어에서 높은 비중을 차지함으로 다른 단어에 비해서 상대적으로 가중치가 크게 나타남.
- **TF-IDF 워드 클라우드**
  - TF-IDF의 경우 특정 문서 내에서 단어 빈도가 높을 수록, 전체 문서들 중 그 단어를 포함한 문서가 적을 수록 TF-IDF값이 높아짐.
  - *보조금, 행안부, 지자체*와 같이 모든 문서에서 흔하게 나타나지 않지만 특정 문서에서 빈도가 높게 나타나는 단어들의 가중치가 크게 나타남.

Fig. 빈도 워드 클라우드

Fig. TF-IDF 워드 클라우드

토픽 모델링

LDA

- **토픽 모델링**
  - LDA방법론을 이용해서 토픽 모델링을 수행함.
  - LDA(Latent Dirichlet Allocation)은 토픽 모델링 방법 중 하나로, 대규모 텍스트 문서 집합에서 어떤 주제들이 존재하는지를 발견하고, 각 문서가 이러한 주제들로 어떻게 구성되어 있는지를 추론하는 확률적 모델임.
- **단어 벡터화**
  - LDA를 수행하기 위해서는 먼저 단어를 벡터로 만들어야 하며 이를 위한 방법으로는 주로 BoW(Bag of Words)와 TF-IDF방법론이 있음..
    - BoW: 텍스트를 토큰(단어 또는 문자)으로 분할하고, 토큰들의 출현 빈도를 기반으로 텍스트를 수치적인 벡터로 표현함.
    - TF-IDF: 특정 문서에서 특정 단어의 중요도를 측정하여 문서들 간의 유사성을 파악하는데 도움이 됨.
  - LDA의 동작 방식 자체가 토픽 분포와 단어 빈도를 기반으로 주어진 단어가 어떤 주제에 속하는지를 추정하기 때문에 빈도정보를 그대로 벡터화 과정에 반영한 BoW를 일반적으로 사용함.
  - 하지만, 데이터와 목적 및 결과에 따라 적절한 방법을 선택해야 하며 본 과정에서는 BoW와 Tf-Idf를 모두 활용하여 LDA 토픽 모델링을 수행함.

토픽 모델링

LDA

- **최적 토픽 수 지정 방법**
  - LDA토픽 모델링에서 토픽 수는 하이퍼 파라미터로 문서가 담고 있는 최적의 값을 직접 지정해주어야 함.
  - Coherence점수는 토픽이 얼마나 의미론적으로 일관성 있는지 판단하는 지표로, 높을수록 의미론적 일관성이 높음. 이를 통해서 해당 모델이 얼마나 실제로 의미 있는 결과를 내는지 확인할 수 있음.
  - Coherence 점수를 통해서 최적의 토픽 수를 지정함.

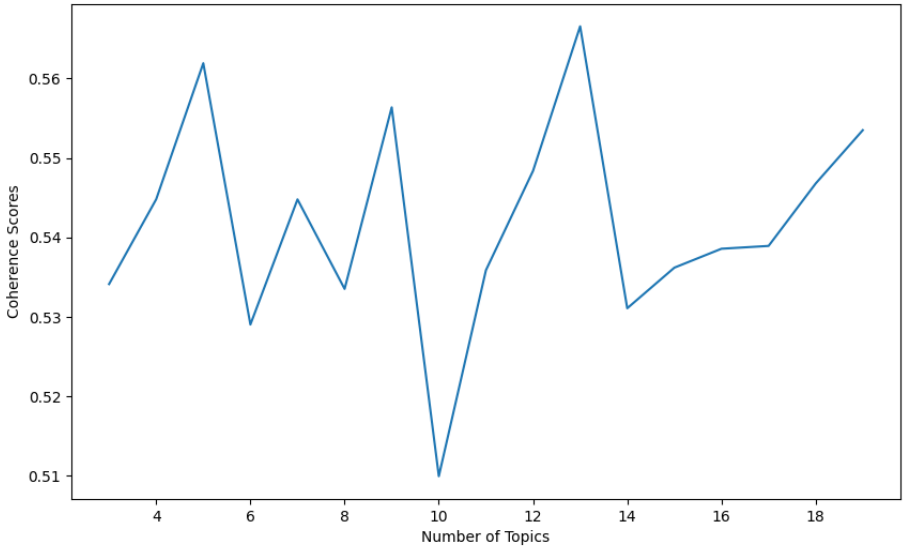


Fig. BoW Coherence Score

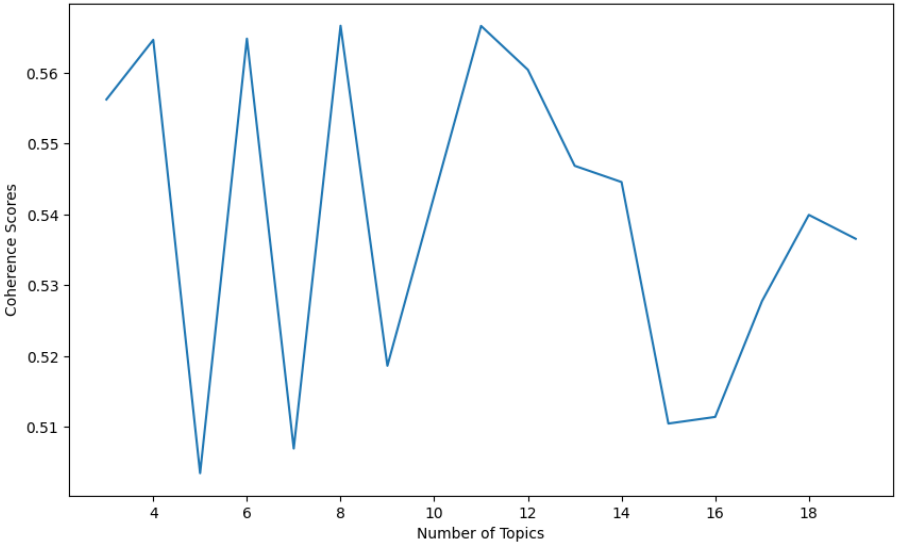


Fig. TF-IDF Coherence Score

01

AI &
HEALTHCARE LAB

토픽 모델링

LDA_BoW

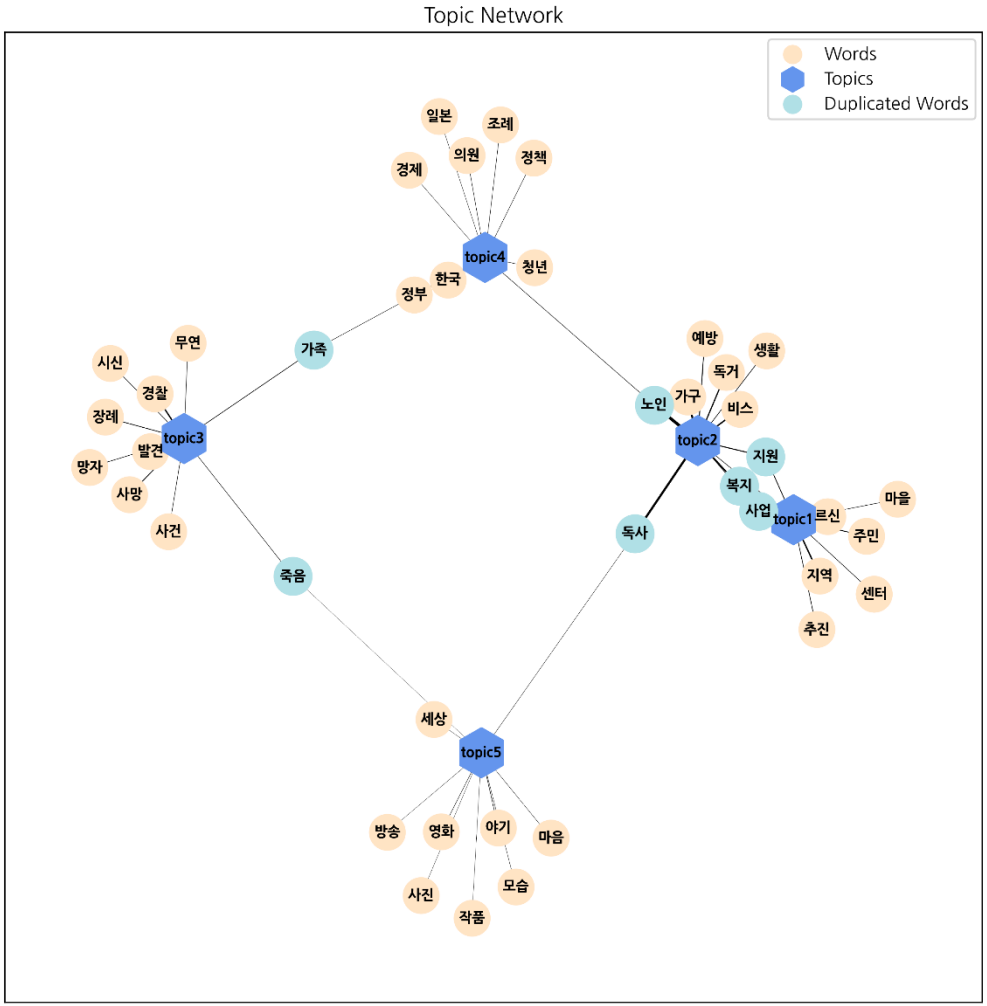
- **토픽 수**
  - Coherence Score로 판단한 최적의 토픽 수 : 5
- **토픽 별 주요 단어 및 연관 확률**

	토픽1	토픽2	토픽3	토픽4	토픽5
키워드1	사업(0.029)	노인(0.042)	발견(0.025)	노인(0.007)	이야기(0.007)
키워드2	지역(0.020)	고독사(0.031)	경찰(0.019)	가족(0.006)	영화(0.006)
키워드3	복지(0.015)	가구(0.028)	사망(0.012)	청년(0.006)	고독사(0.005)
키워드4	주민(0.012)	서비스(0.020)	장례(0.010)	정책(0.006)	마음(0.005)
키워드5	어르신(0.011)	복지(0.019)	가족(0.009)	의원(0.006)	모습(0.005)
키워드6	지원(0.011)	독거(0.017)	시신(0.008)	의원(0.006)	작품(0.004)
키워드7	추진(0.009)	지원(0.014)	사망자(0.008)	조례(0.006)	사진(0.004)
키워드8	센터(0.008)	예방(0.012)	사건(0.007)	일본(0.005)	방송(0.004)
키워드9	노인(0.008)	사업(0.010)	무연(0.007)	정부(0.005)	세상(0.004)
키워드10	마을(0.007)	생활(0.009)	죽음(0.007)	한국(0.005)	죽음(0.003)

토픽 모델링

LDA_BoW

- **토픽 네트워크 시각화**
  - Duplicated Words는 2개 이상 토픽에서 중복되어서 나타난 단어를 말함.



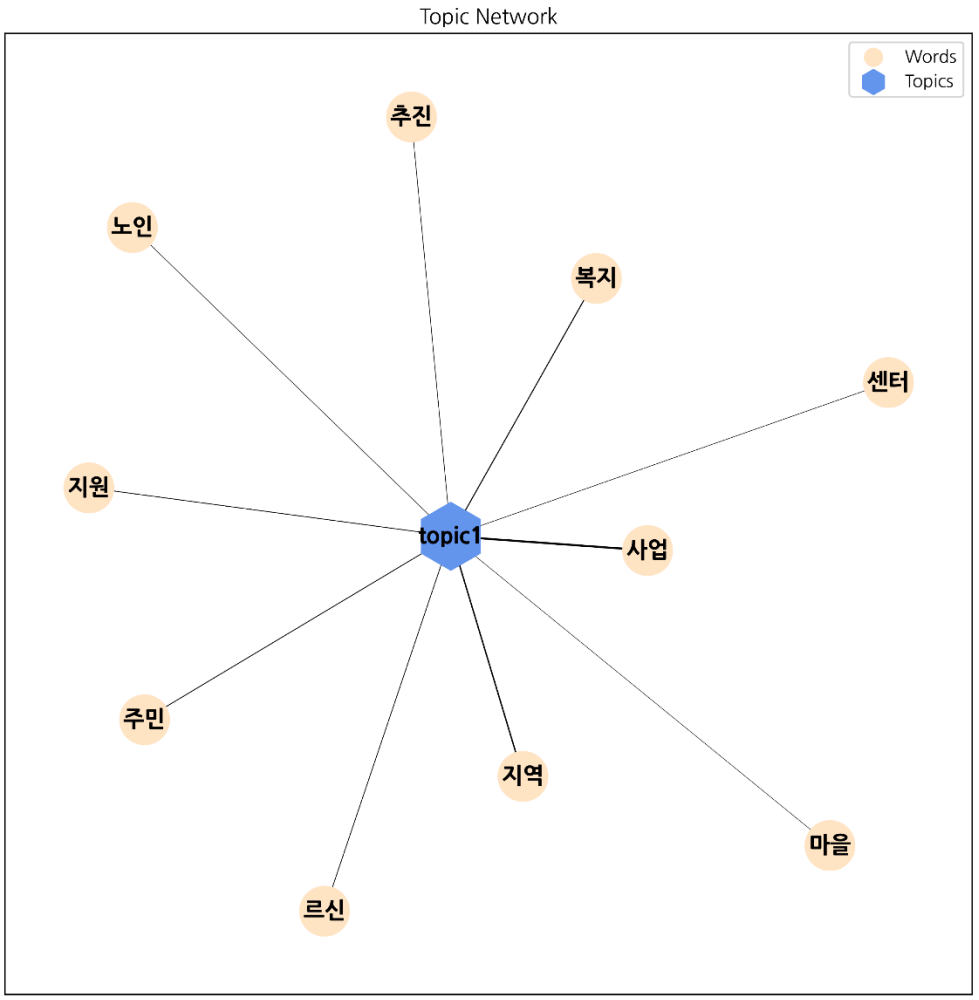
01

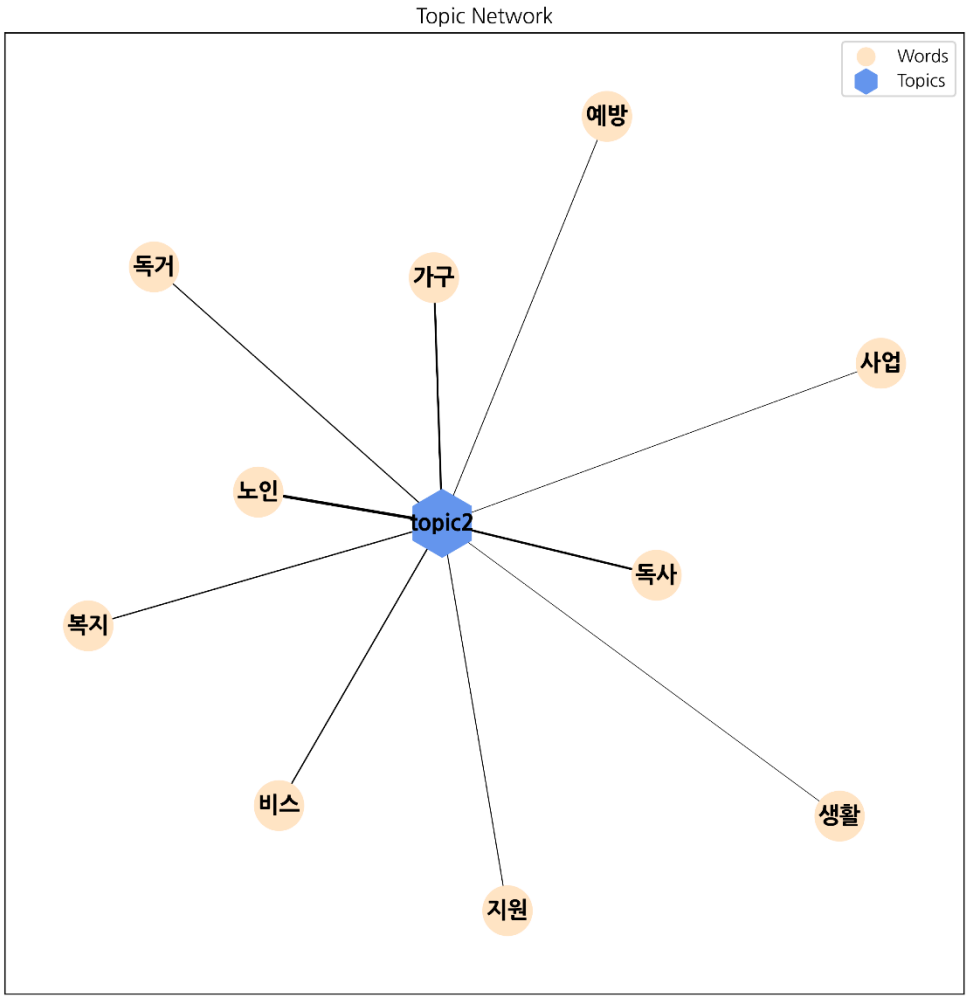
AI &
HEALTHCARE LAB

토픽 모델링

LDA_BoW

- 각 토픽 별 네트워크 시각화 : Topic1, Topic2

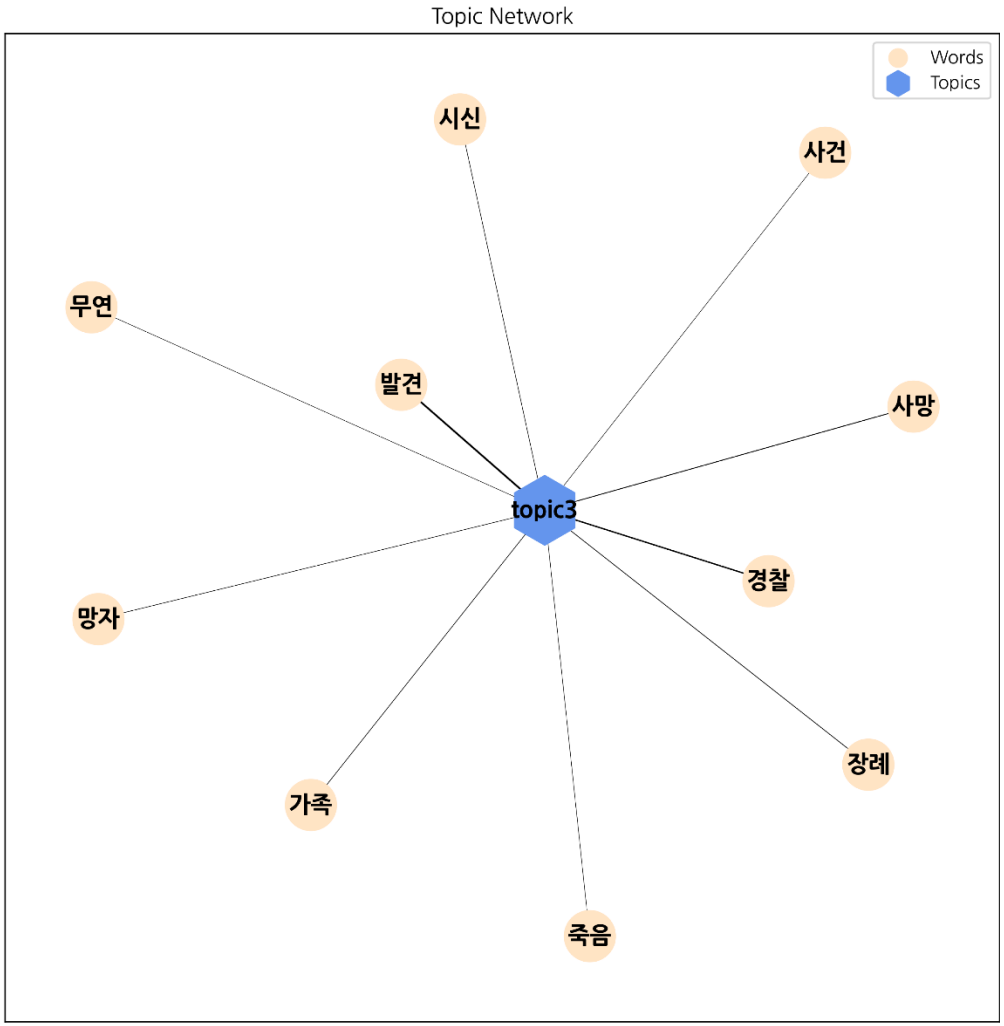


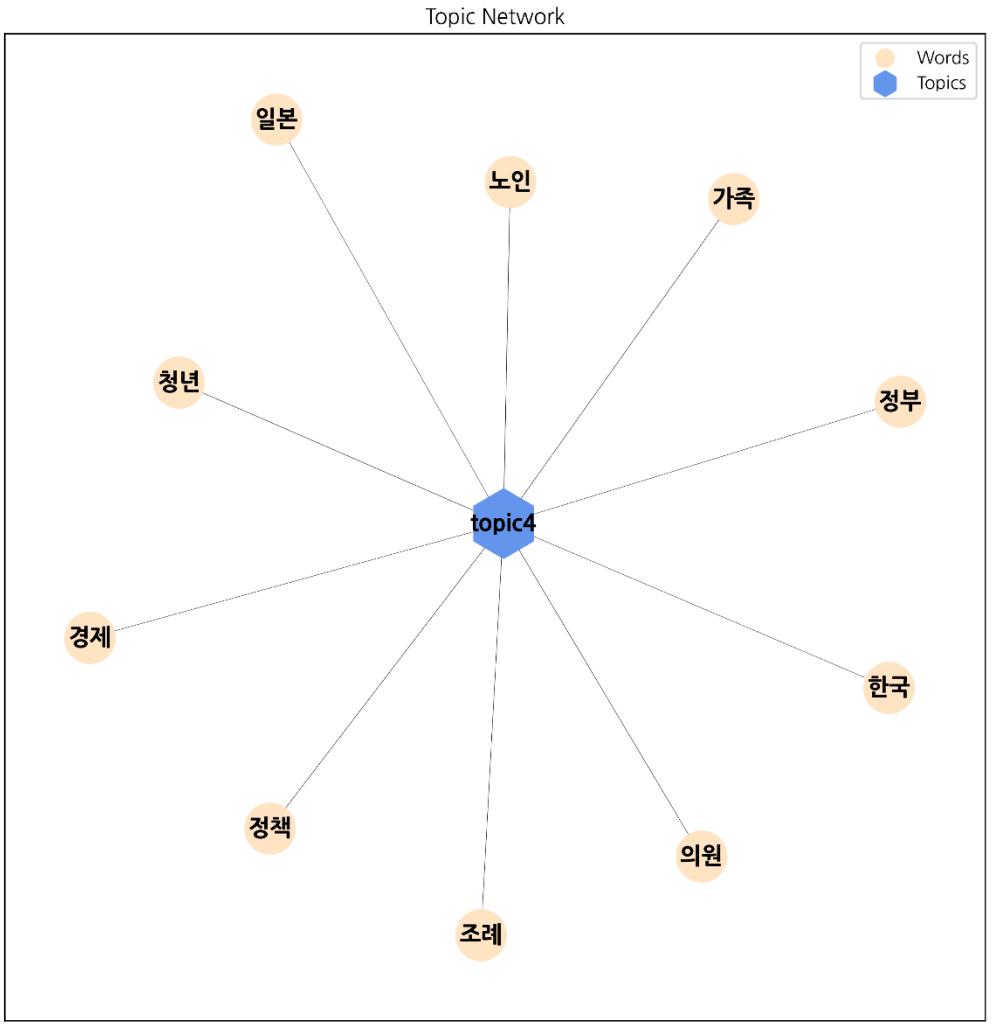


토픽 모델링

LDA_BoW

- 각 토픽 별 네트워크 시각화: Topic3, Topic4

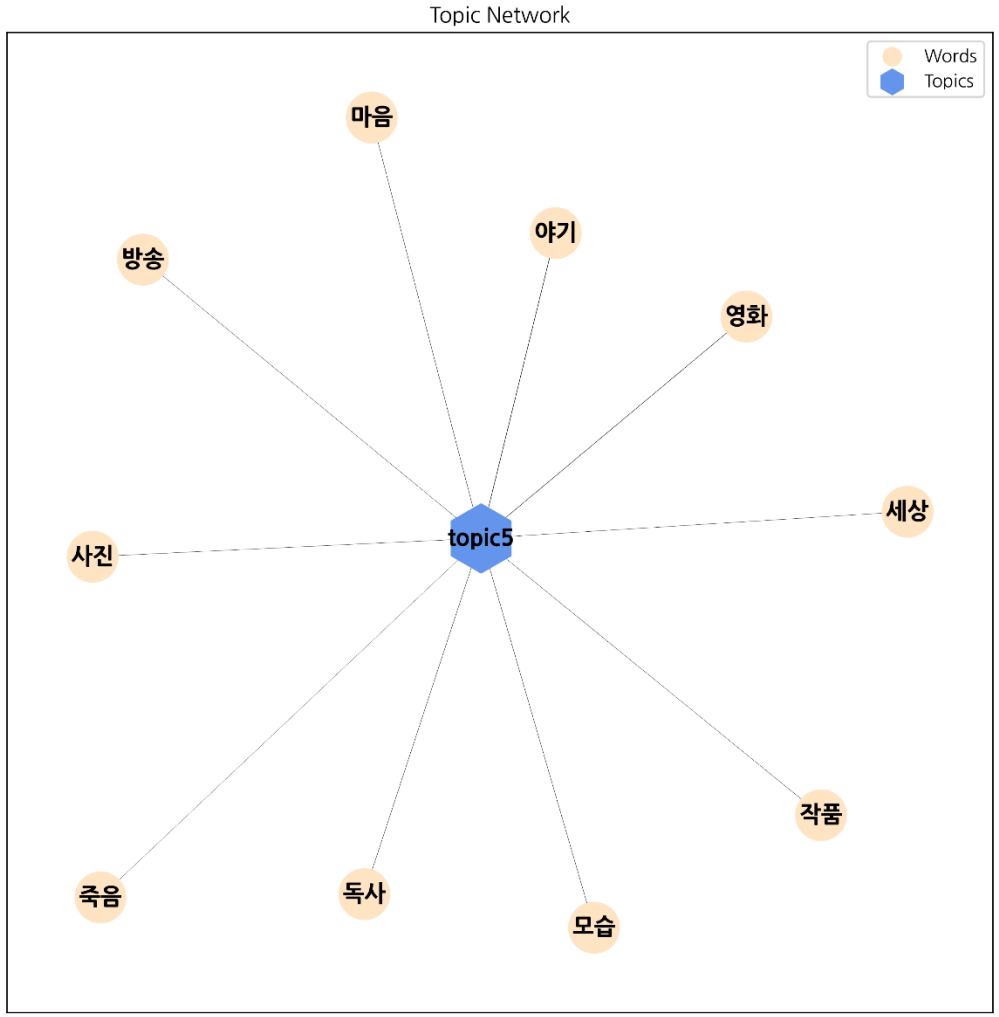




토픽 모델링

LDA_BoW

- 각 토픽 별 네트워크 시각화: Topic5



01

AI &
HEALTHCARE LAB

토픽 모델링

LDA_TF-IDF

- **토픽 수**
  - Coherence Score로 판단한 최적의 토픽 수 : 4
- **토픽 별 주요 단어 및 연관 확률(표)**

	토픽1	토픽2	토픽3	토픽4
키워드1	조례(0.004)	안양시(0.002)	노인(0.006)	노인(0.005)
키워드2	경찰(0.004)	보령시(0.001)	사업(0.005)	사망자(0.002)
키워드3	의회(0.003)	광양시(0.001)	독거(0.005)	죽음(0.002)
키워드4	발견(0.003)	안양시장(0.001)	서비스(0.004)	무연(0.002)
키워드5	의원(0.003)	광양제철소(0.001)	복지(0.004)	장례(0.002)
키워드6	발의(0.002)	김해시(0.001)	어르신(0.004)	가족(0.002)
키워드7	의정(0.002)	성매매(0.001)	가구(0.004)	가구(0.002)
키워드8	임시회(0.001)	과천시(0.001)	홀몸(0.003)	증가(0.002)
키워드9	신고(0.001)	비산(0.001)	지역(0.003)	청주시(0.002)
키워드10	의장(0.001)	터전(0.001)	지원(0.003)	독거(0.002)

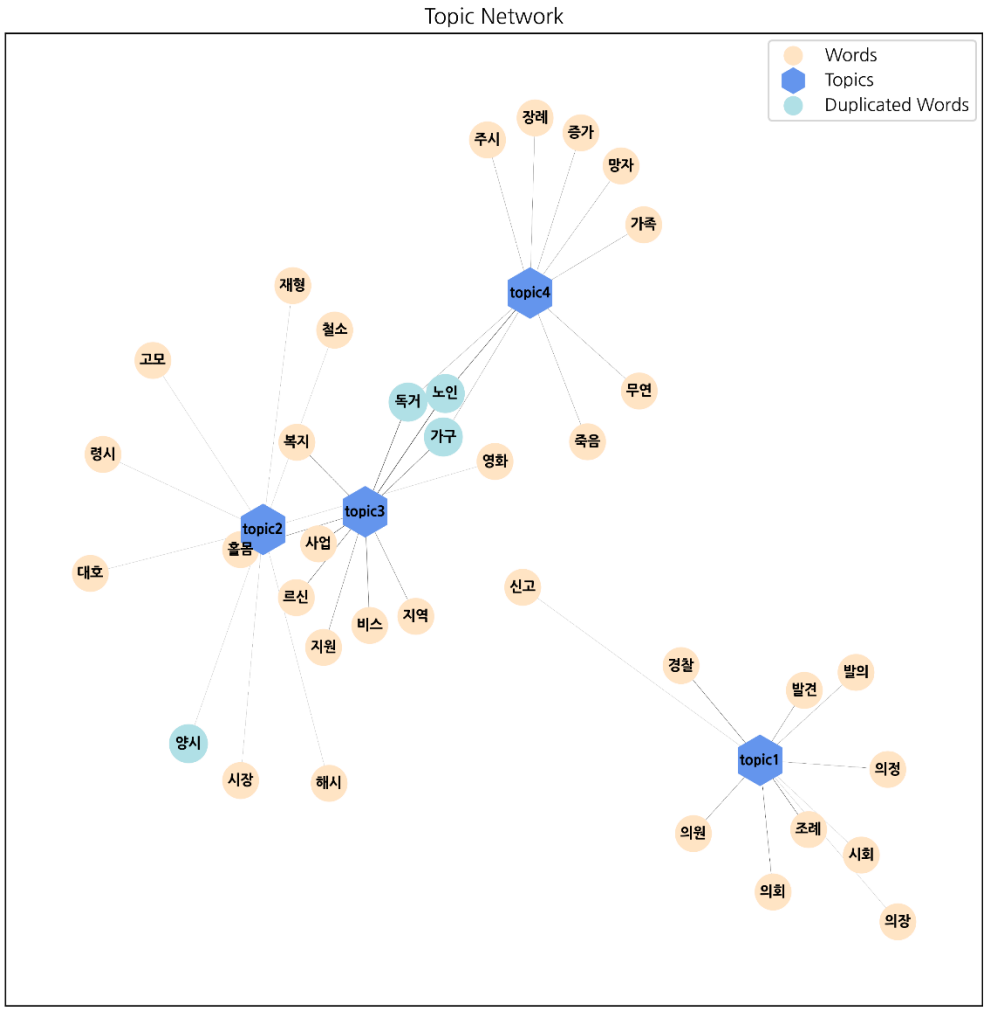
01

AI &
HEALTHCARE LAB

토픽 모델링

LDA_TF-IDF

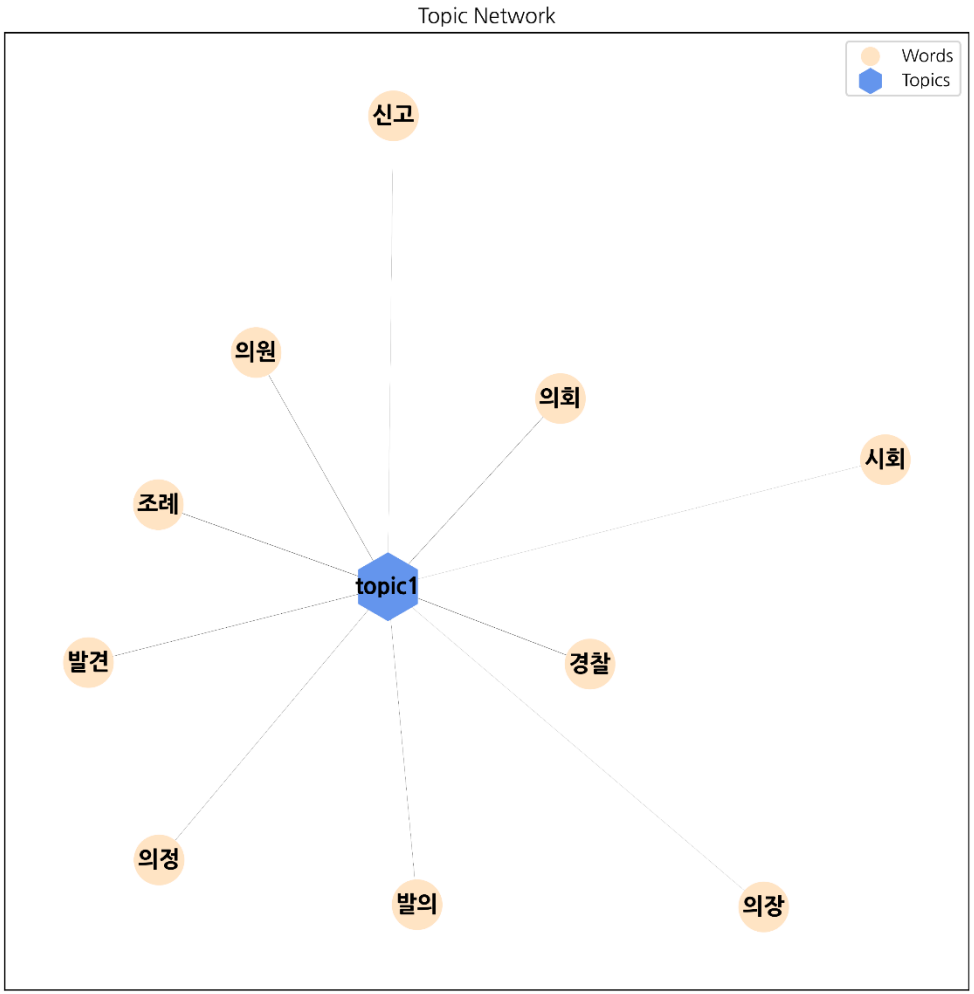
- **토픽 네트워크 시각화**
  - Duplicated Words는 2개 이상 토픽에서 중복되어서 나타난 단어를 말함.

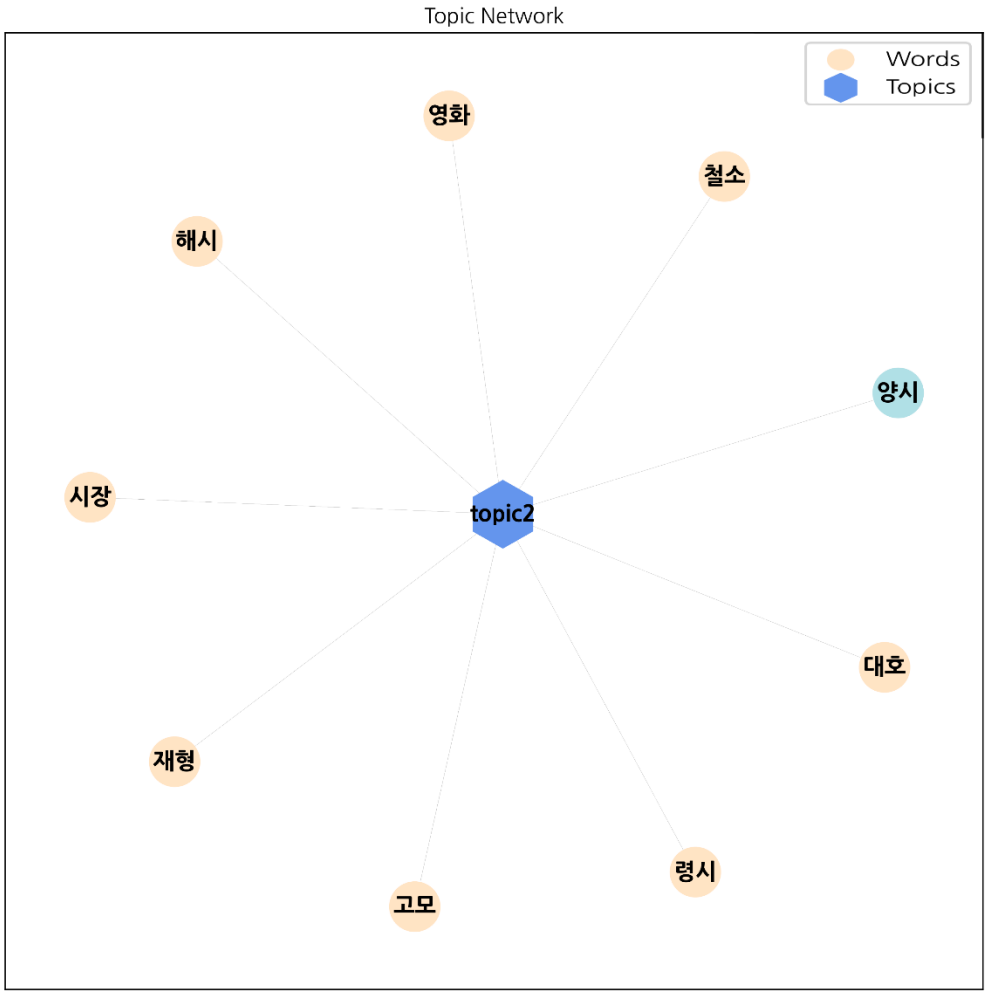


토픽 모델링

LDA_TF-IDF

- 각 토픽 별 네트워크 시각화: Topic1, Topic2

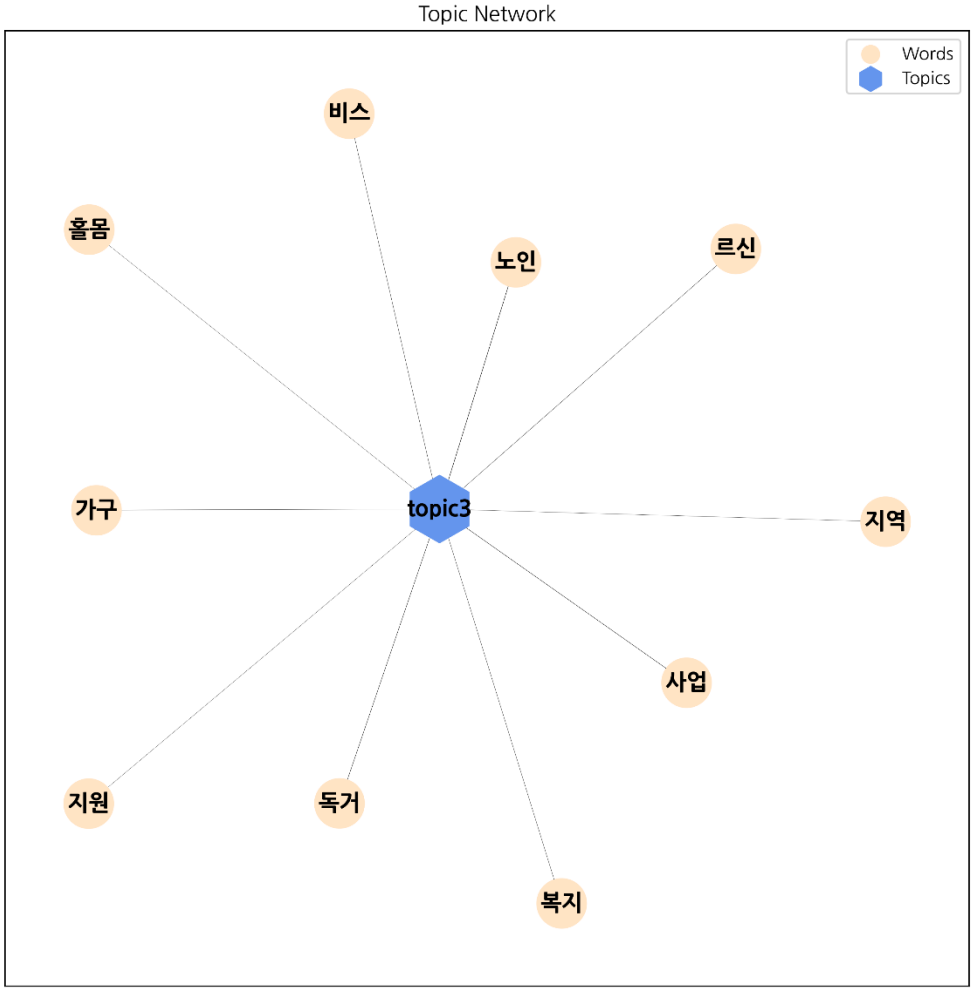


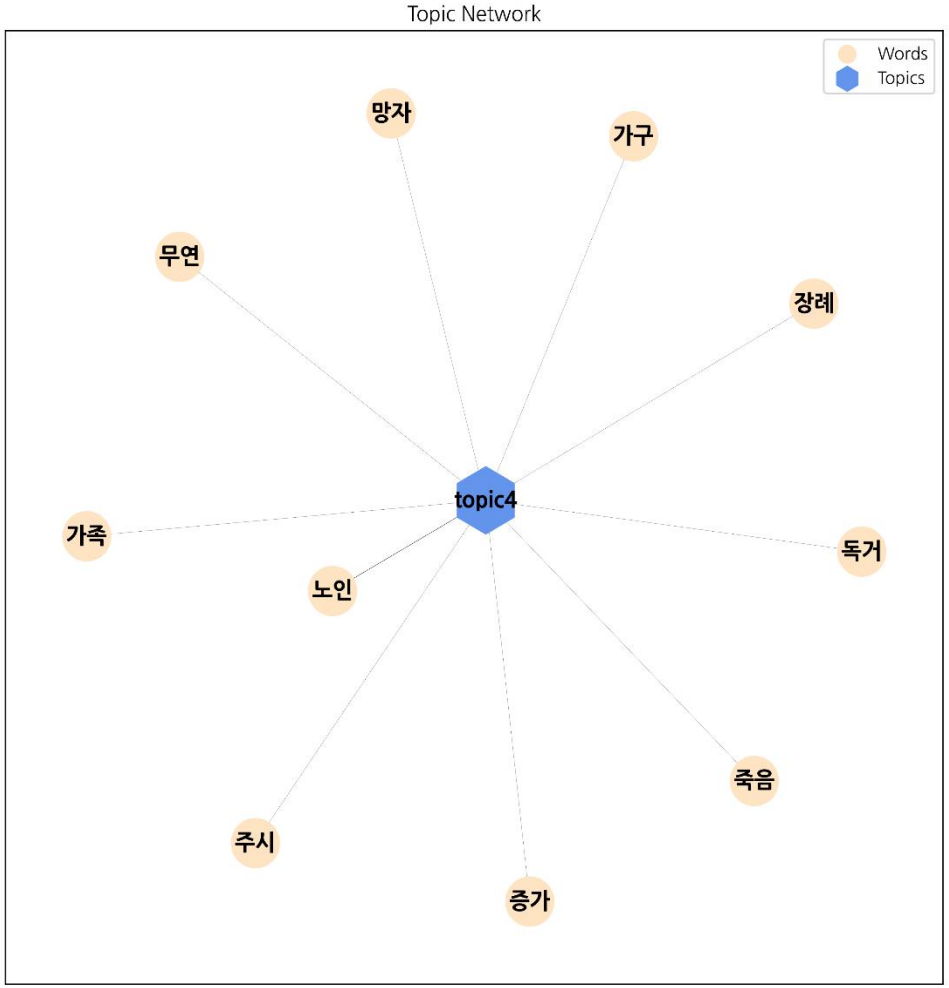


토픽 모델링

LDA_TF-IDF

- 각 토픽 별 네트워크 시각화: Topic3, Topic4





토픽 모델링

토픽별 뉴스기사

- **토픽에 해당하는 주요 기사의 원문**
  - 각 토픽 별로 주요 기사의 요약문을 예시로 살펴봄
  - Topic1
    - ✓ *정부와 지자체가 비영리 민간단체에 대해 '1단체에 1사업' 원칙을 어기고 동일한 장소에서 6개의 사업에 중복으로 보조금을 지원하고 있다. 또한, 예산이 남는 경우에도 같은 사업에 추가 지원이 이뤄지며, 반환처분 받은 곳에 다시 지원도 이루어지고 있다. 정부는 하반기에 4000건 이상의 사업에 대해 집중 점검을 실시할 계획이다.*
    - ✓ *희망브리지 회장 송필호는 재난 맞춤형 지원을 강조하며, 취약 계층의 회복을 최선으로 노력해야 한다고 말했습니다. 희망브리지는 삶의 전반에서 다양한 재난에 대응하여 맞춤형 지원을 목표로 하고 있으며, 지역공동체를 강화하는 방안을 모색하고 있습니다. 회장은 또한 민간단체가 유연하고 창의적인 방법으로 어려운 사람들을 돌봄으로써 사회구조적 문제를 고민해야 한다고 강조했습니다.*
    - ✓ 경기도지사 김동연은 과장급 워크숍에서 '도전(Try)', '열정(Energy)', '꿈(Dream)'을 주제로 정책토론을 벌이며 유쾌한 반란을 일으키고자 했다고 강조했습니다. 워크숍은 경기도 과장급과 공공기관 경영본부장급 280여 명이 참석하여 정책 발굴과 논의를 진행한 것으로, 우수한 아이디어 20개를 선택해 참가자들의 현장 투표와 심사위원단 평가를 통해 선정되었습니다. 경기도는 이와 더불어 '2023 기회경기 팀장급 공감 워크숍'을 예정하고 있습니다.

토픽 모델링

토픽별 뉴스기사

- **토픽에 해당하는 주요 기사의 원문**
  - 각 토픽 별로 주요 기사의 요약문을 예시로 살펴봄
  - Topic2
    - ✓ *50대 지체 장애인이 혼자 사는 공공임대주택에서 두 달 만에 발견되었다. 경찰은 집안에서 사망한 장애인을 발견하고, 유서에는 화장 후 유골을 산에 뿌리고 장례비로 돈을 사용해달라는 내용이 담겨 있었다. 경찰은 타살 혐의가 없어 단순 변사로 마무리했으며, 용인시는 무연고 공영장례를 치렀다. 또한, 시에서는 취약계층에 대해 전화와 가정방문을 실시하고 있으며 고립가구에 대한 세심한 관심을 가지고 대응할 예정이다.*
    - ✓ *용인시의 한 공공임대주택에서 혼자 사는 50대 지체 장애인이 두 달 만에 사망한 것이 발견되었다. A 씨는 지체 장애를 가진 채 혼자 살았으며, 현장에서는 극단적 선택에 사용된 것으로 보이는 물건과 두 달 전 작성한 유서가 발견되었다. 경찰은 타살 혐의가 없어 단순 변사로 처리하며, 용인시는 취약계층에 대해 세심한 관심을 가지고 대응할 예정이다.*
    - ✓ *어느 사람의 친형이 7년 동안 집에서 나오지 않는 은둔형 외톨이로 살고 있다는 사연이 올라왔습니다. 전문가들은 해당 개인의 정신기저질환과 조현병 가능성을 언급하며 정신병원 입원 및 치료를 권장하고 있습니다. 가족들은 입원적합성심사위원회 통과를 통해 입원을 시도하고 있으며, 해당 개인은 심각한 상태로 상태가 점점 악화되고 있습니다.*

토픽 모델링

토픽별 뉴스기사

- **토픽에 해당하는 주요 기사의 원문**
  - 각 토픽 별로 주요 기사의 요약문을 예시로 살펴봄
  - Topic3
    - ✓ *동대문구청과 hy가 협약하여 '고독사 위험가구 건강음료 지원사업'을 추진한다. 기초생활수급자와 차상위 계층을 대상으로 건강음료를 전달하고 안부 확인을 실시하며, 이는 1억 원으로 개별 지자체와 맺은 협약 중 가장 큰 규모다. hy는 자사 유통망인 '프레시 매니저'를 활용해 대상자 1061명에게 건강음료를 전달하고 위기 상황 예방을 위해 동대문구청과 실시간 소통을 한다고 한다.*
    - ✓ *구미시 인동동에 홀로 거주하는 78세의 이 모 씨가 인공지능(AI) 스피커를 이용하여 응급 상황에서 구조되었다. AI 스피커가 SOS 신호를 보내면서 119 구급대원들이 신속하게 출동해 이 씨를 응급 이송했다. 구미시는 취약계층 1인 가구에 AI 스피커를 설치하여 인공지능 통합 돌봄 서비스를 시작했으며, 위급한 상황 시 돌봄 대상자를 구조하는 긴급호출 서비스뿐만 아니라 말벗 역할도 수행한다고 한다.*
    - ✓ *서대문구가 명예사회복지공무원으로 이뤄진 '복지사각지대 발굴단'을 통해 위기가구를 적극 발굴하고 지원할 예정이며, 올 연말까지 명예사회복지공무원을 3300여 명까지 확대할 계획이라고 밝혔습니다. 또한 구는 다양한 방법을 통해 위기가구를 찾아낼 예정이라고 설명했습니다.*

토픽 모델링

토픽별 뉴스기사

- **토픽에 해당하는 주요 기사의 원문**
  - 각 토픽 별로 주요 기사의 요약문을 예시로 살펴봄
  - Topic4
    - ✓ *미국 작가 에드워드 호퍼의 작품 '푸른 저녁'은 프랑스에서 그렸으며, 독특한 점은 7명의 인물이 등장하지만 서로 눈을 마주치지 않고 대화를 하지 않는 것이다. 이 그림은 호퍼가 야심 차게 준비한 큰 작품으로, 미국 작가가 프랑스에서 그린 것 중 가장 큰 축에 속한다. '푸른 저녁'은 대중적인 반응을 얻지 못해 평생 전시되지 않았으며, 후에 미국의 모습에 집중하는 작품들로 인기를 얻게 되었다.*
    - ✓ *에드워드 호퍼의 대표작인 '푸른 저녁'은 프랑스에서 그린 작품으로, 1914년에 그렸습니다. 그림은 영어가 아닌 프랑스어로 'Soir Bleu'라는 제목을 가지고 있으며, 어두운 옷을 입은 남성들 가운데 분칠을 한 피에로와 여성이 눈길을 사로잡고 있습니다. 이 그림은 호퍼의 야심작으로 그렸으며, 그림 속 인물들은 모두 서로 눈을 맞추거나 대화를 나누지 않고 고독한 모습을 보여주고 있습니다. 이 그림은 처음 전시했을 때 큰 반응을 얻지 못해 주목받지 못했으나, 뒤이어 그가 그린 다른 작품들과 함께 인정받아 성공적인 예술가로 자리매김하게 됩니다. 이후 호퍼는 미국의 일상을 그리는 작품으로 미국 미술사의 중요한 작가로 인정받게 되었습니다.*
    - ✓ *60세로 세상을 떠난 농구 스타 출신 김영희 씨. 1980년대 한국 여자농구 전성기를 빛냈고, 1982년과 1986년에는 인도 뉴델리와 서울 아시아경기에서 센터로 뛰며 은메달을 획득했습니다. 그 후 뇌종양으로 인한 거인병과 싸우며 60세에 별세했습니다. 마르판 증후군과 싸우던 농구 스타 한기범 씨도 두 차례 심장 수술을 받았습니다. 김영희 씨는 다양한 재난에 시달리며, 자선 활동을 하며 남에게 베푸는 삶을 살았습니다.*

토픽 모델링

토픽별 뉴스기사

- **토픽에 해당하는 주요 기사의 원문**
  - 각 토픽 별로 주요 기사의 요약문을 예시로 살펴봄
  - Topic5
    - ✓ *우리나라의 고독사 위험군은 153만 명으로, 전체 1인 가구의 1명 중 5명이 해당된다. 2021년에는 3378명의 고독사가 발생하여 사회적인 문제로 인식되고 있다. 고독사 위험은 고령층이 아닌, 경제적으로 취약한 40~60대 중장년층에서 더 많이 발생하며, 정부는 이러한 위험군을 찾아내고 지원하는 사회안전망 구축이 시급하다고 강조하고 있다.*
    - ✓ *우리나라의 고독사 위험군은 153만 명으로, 1인 가구 5명 중 1명이 해당된다. 2021년 기준으로 1.06명씩 고독사가 발생하는데, 정부는 2027년까지 이 수를 20% 줄여 0.85명으로 감소시키는 것을 목표로 한다. 중장년층 남성은 은퇴 후에 경제적으로 좌절하고 고립되는 경우가 많아 고독사 위험에 처할 가능성이 크다. 정부는 이웃들이 고독사 위험군을 발굴해 전문가에게 의뢰하는 '우리 마을 지킴이'를 양성하여 복지 서비스로 연계하려고 한다.*
    - ✓ *정부는 사회적 고립으로 인해 홀로 사망하는 '고독사' 사망자 수를 2027년까지 현재보다 20% 줄이기 위해 기본계획을 발표했다. 기본계획에는 고독사 위험군을 발굴해 위기요인을 해소하기 위한 복지 프로그램을 연계 지원하고, 정보통신(IT) 기술을 활용해 주거지 내 위기 징후를 조기에 포착할 방침이 담겼다. 복지부는 이를 수행하는데 소요되는 예산규모를 잠정 3907억원으로 추산했다.*