

합격만 주소서

20181480 권도윤

20181483 박명석

20171468 서민지



목차



01

문제 정의

02

데이터 수집

03

데이터 전처리

04

분석 및 모델링

05

기대효과 및 한계점

인사담당자들은 '소제목으로 내용을 정리한 형태의 자기소개서(38.9%)'를 가장 선호한다고 답했다. 이 외에도 △인터뷰 기사와 같이 개성있는 문체로 작성된 자기소개서(27.0%) △회사 양식에 맞춰 잘 정리된 형태(26.2%)의 자기소개서 순으로 선호한다고 답했다.

출처 : 소비라이프

"올 추석에는 자소서나 쓸래요"...명절에도 고향 못가는 취준생

출처 : 조선일보

구직자, 88.1% 자소서포비아 겪어

※ 신입직 구직자 2,158명 대상 조사, 자료: 잡코리아X알바몬

Q. 자소서 작성에 어려움을 느끼는 이유(복수응답)

1위 무엇을 적어야 할지 막막하다, 59.9%

2위 어떤 부분을 강조해야 할지 어렵다, 47.8%

3위 그럴 듯한 스토리가 부족하다, 33.8%

4위 업적을 보여주는 데 한계가 있다, 24.6%

5위 제한된 글자 수를 맞추기 어렵다, 13.5%

군대

학업

출신학교

영어



크롤링 패키지 - BeautifulSoup

JOBKOREA

데이터 개수

25,574

수집 목록

질문, 답변, 조언, 스펙, 평가, 총평, 주소,
회사명, 지원시기, 근무형태, 직무분야

saramin

데이터 개수

16,902

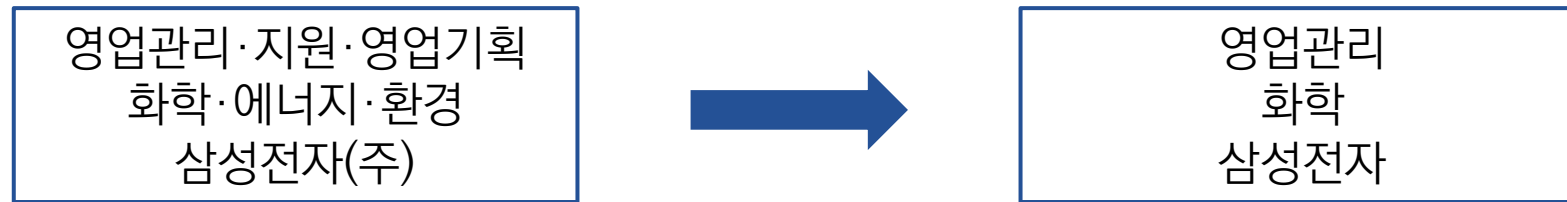
수집 목록

질문, 답변

코리아세븐의 영업관리 직무 수행에 있어 중요한 역량은 데이터 분석 능력과 소통 능력입니다. 정확한 상권 분석과 매출 트렌드 분석을 통하여 시즌에 맞고 고객들이 필요로 하는 트렌디한 상품들을 본사의 정책과 각각의 점포의 상황에 맞게 적용해야 하기 때문입니다. 직접 다양한 데이터를 수집 분석함과 동시에 지속적인 소통을 통해 효율적으로 프로모션을 진행하여 저의 역량을 발전시켜 나가면서도 좋은 성과를 얻은 경험이 있습니다. 에서 음료 부문 담당을 맡게 되며 3개월 이내에 전임 관리자의 매출보다 10 이상 성장시키겠다는 목표를 수립하고 목표 달성을 위해 데이터 분석을 하며 파트너사와의 협업을 통한 프로모션을 준비하였습니다. 소비 트렌드를 읽어내기 위해 각종 매체를 이용해 트렌드 관련 정보를 수집하고 타 매장과 차별성을 두기 위해 직접 시장조사 및 분석을 하며 정확한 데이터 분석 능력을 키웠습니다. 이러한 데이터 분석과 동시에 롯데칠성 제과와 함께 1개월간의 지속적인 소통을 통해 발생 가능한 문제점을 제거하고 효율적인 진행 방안을 수립하여 나들이 철 롯데 음료 과자 모음전 프로모션을 진행하였고 그 결과 총매출 5 성장과 타사 제품대비 프로모션 진행 제품 매출 10 성장이라는 성과를 얻을 수 있었습니다. 이렇게 업무에 필요한 역량이 무엇인지 정확히 파악하고 그 역량을 발전시키기 위해 노력하여 좋은 성과를 이뤄낸 경험과 이 경험을 통해 얻은 데이터 분석 능력 소통 능력을 업무 수행에 강점으로 활용하여 코리아세븐의 성장에 기여하겠습니다'

- 소제목 추출
- 특수기호 및 줄바꿈 기호 제거
- 좋은점, 아쉬운점 평가 부분 제거

> 텍스트 전처리는 주로 re 패키지를 사용해 정규식 패턴을 정의하여 처리했습니다.



- 직무분야 분리
- 회사명 처리

> 텍스트 전처리는 주로 re 패키지를 사용해 정규식 패턴을 정의하여 처리했습니다.

특정 분야 및 기업에 관한 키워드 제공



응용프로그래머



빅데이터·AI(인공지능)



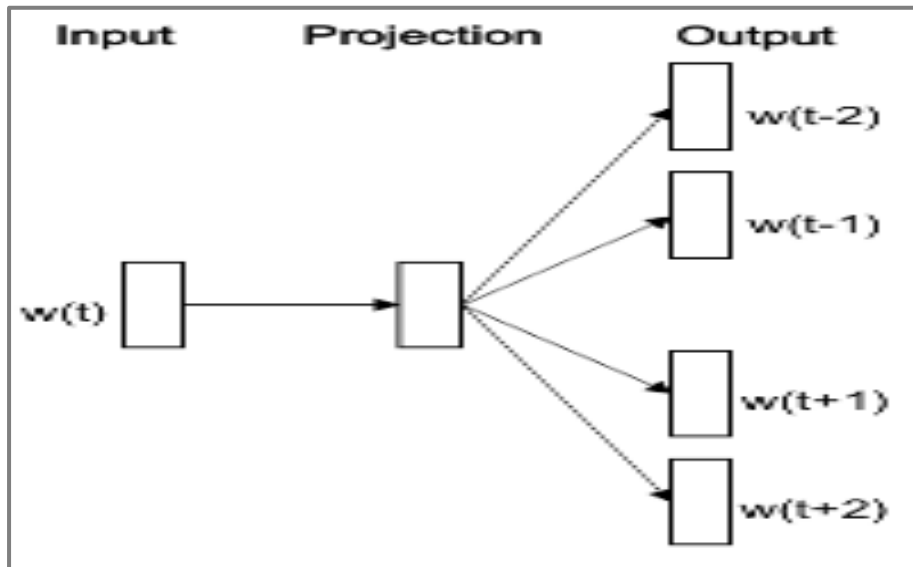
DBA·데이터베이스



단어 추천 제공

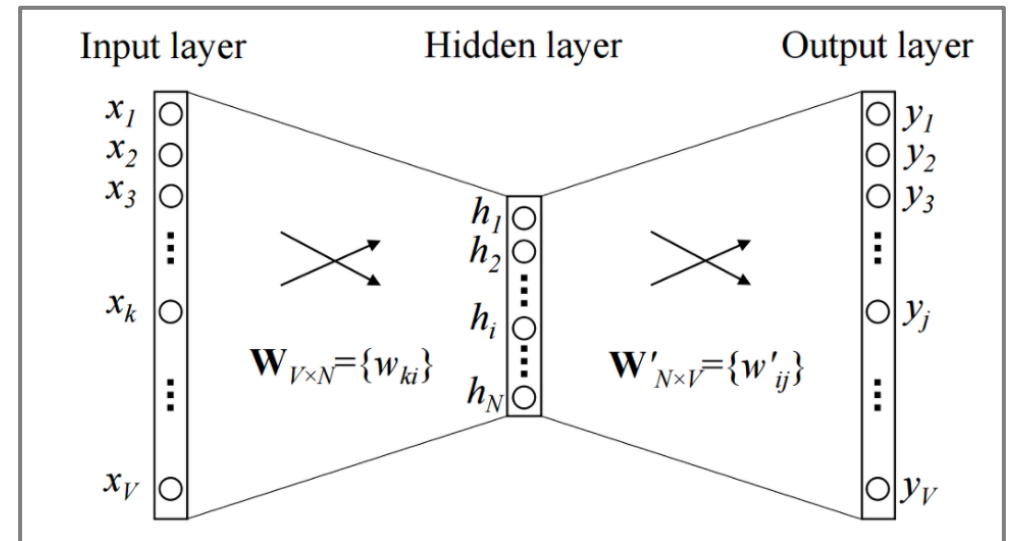
Word2Vec

단어 간 유사도를 반영할 수 있도록
단어의 의미를 벡터화 할 수 있는 대표적인 방법



FastText

Word2Vec의 확장판, Word2Vec과 다르게 등장 빈도 수가 적은 단어라 하더라도 n-gram으로 임베딩을 하는 특성상
참고할 수 있는 경우의 수가 많아져 정확도가 높음
하나의 단어 안에도 여러 단어들이 존재하는 것으로 간주
(내부 단어를 고려하여 학습)



단어 추천 제공

Word2Vec

키워드 : 노력	키워드 : 결정
자기계발	판단
정진	선택
헌신	내리다
총력	신중하다
배움	후회
심혈	의사결정
자기개발	결론

FastText

키워드 : 노력	키워드 : 결정
부단하다	판단
담금질	내리다
꾸준하다	선택
끈임	의사결정
자기계발	취약성
자기개발	결정성
거북이	좌지우지

중요 문장 제공

TextRank

word graph 나 sentence graph 를 구축한 뒤,
Graph ranking 알고리즘인 PageRank 를 이용하여
각각 키워드와 핵심 문장을 선택
이들을 이용하여 주어진 문서 집합을 요약

원본문서

전처리

TF-IDF

Graph 생성

TextRank

중요 문장 제공

"역량 발전을 위한 노력, 데이터 분석능력, 소통능력"

코리아세븐의 영업관리 직무 수행에 있어 중요한 역량은 데이터 분석 능력과 소통 능력입니다. 정확한 상권 분석과 매출 트렌드 분석을 통하여 시즌에 맞고 고객들이 필요로 하는 트렌디한 상품들을 본사의 정책과 각각의 점포의 상황에 맞게 적용해야 하기 때문입니다. 직접 다양한 데이터를 수집 분석함과 동시에 지속적인 소통을 통해 효율적으로 프로모션을 진행하여 저의 역량을 발전시켜 나가면서도 좋은 성과를 얻은 경험이 있습니다. 예서 음료 부문 담당을 맡게 되며 3개월 이내에 전임 관리자의 매출보다 10 이상 성장시키겠다는 목표를 수립하고 목표 달성을 위해 데이터 분석을 하며 파트너사와의 협업을 통한 프로모션을 준비하였습니다. 소비 트렌드를 읽어내기 위해 각종 매체를 이용해 트렌드 관련 정보를 수집하고 타 매장과 차별성을 두기 위해 직접 시장조사 및 분석을 하며 정확한 데이터 분석 능력을 키웠습니다. 이러한 데이터 분석과 동시에 롯데칠성 제과와 함께 1개월간의 지속적인 소통을 통해 발생 가능한 문제점을 제거하고 효율적인 진행 방안을 수립하여 나들이 철 롯데 음료 과자 모음전 프로모션을 진행하였고 그 결과 총매출 5 성장과 타사 제품대비 프로모션 진행 제품 매출 10 성장이라는 성과를 얻을 수 있었습니다. 이렇게 업무에 필요한 역량이 무엇인지 정확히 파악하고 그 역량을 발전시키기 위해 노력하여 좋은 성과를 이뤄낸 경험과 이 경험을 통해 얻은 데이터 분석 능력 소통 능력을 업무 수행에 강점으로 활용하여 코리아세븐의 성장에 기여하겠습니다.

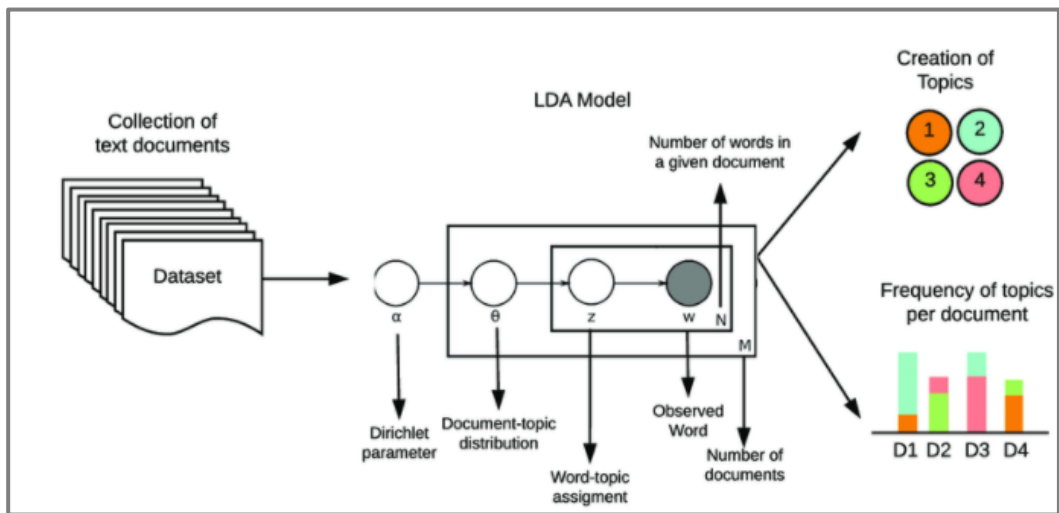
중요 문장

1. 코리아 세븐 의 영업관리 직무 수행에 있어 중요한 **역량**은 **데이터 분석 능력과 소통 능력**입니다.
2. 이렇게 업무에 필요한 **역량**이 무엇인지 정확히 파악하고 그 **역량**을 발전시키기 위해 노력하여 좋은 성과를 이뤄 낸 경험과 이 경험을 통해 얻은 **데이터 분석 능력 소통 능력**을 업무 수행에 강점으로 활용하여 코리아 세븐의 성장에 기여하겠습니다

사용 모형

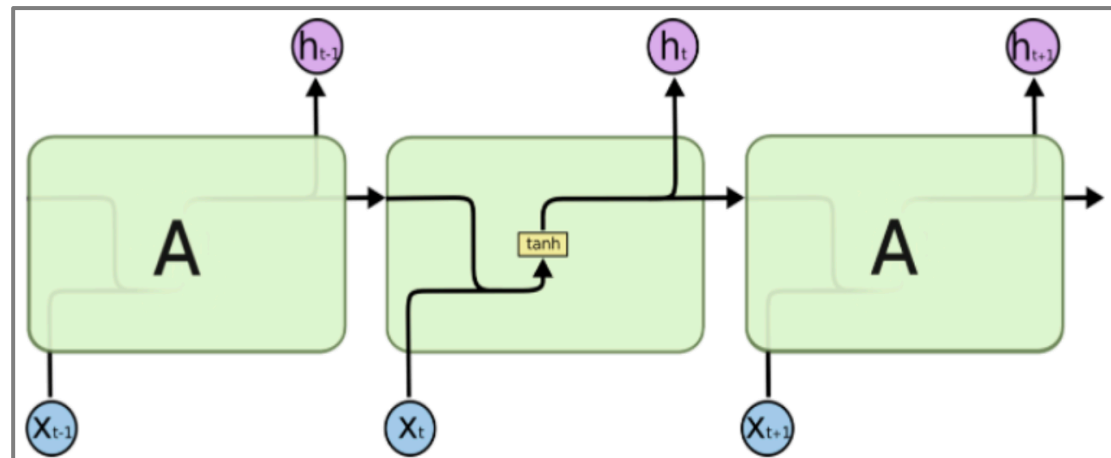
LDA (Latent Dirichlet Allocation)

주제별 단어수 분포를 바탕으로
주어진 문서에서 발견된 단어 수 분포를 분석함으로써
해당 문서가 어떤 주제들을 함께 다루고있을지를 예측



RNN 모형 (Recurrent Neural Network)

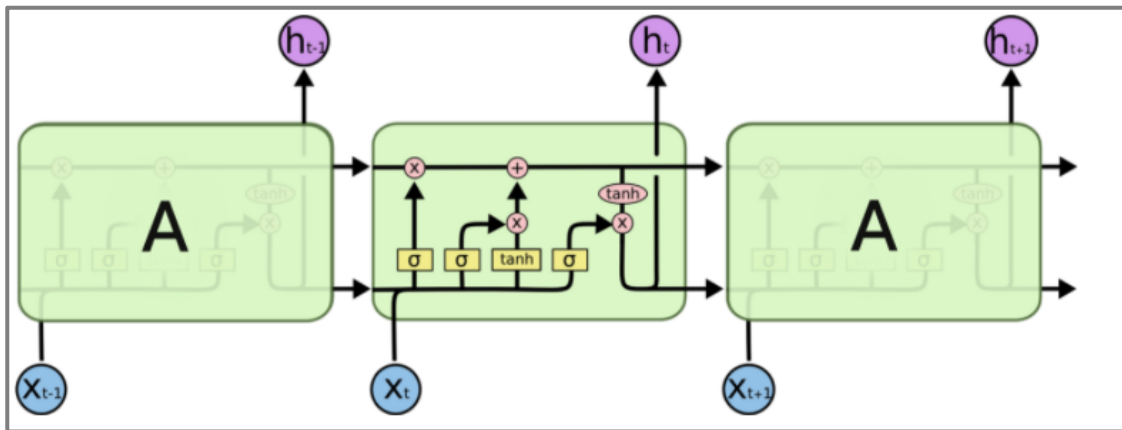
유닛간의 연결이 순환적 구조를 갖는 특징을 가졌고,
이러한 구조는 시변적 동적 특징을 모델링 할 수 있다.
신경망 내부에 상태를 저장할 수 있어 시퀀스 형태 입력을 처리 가능



사용 모형

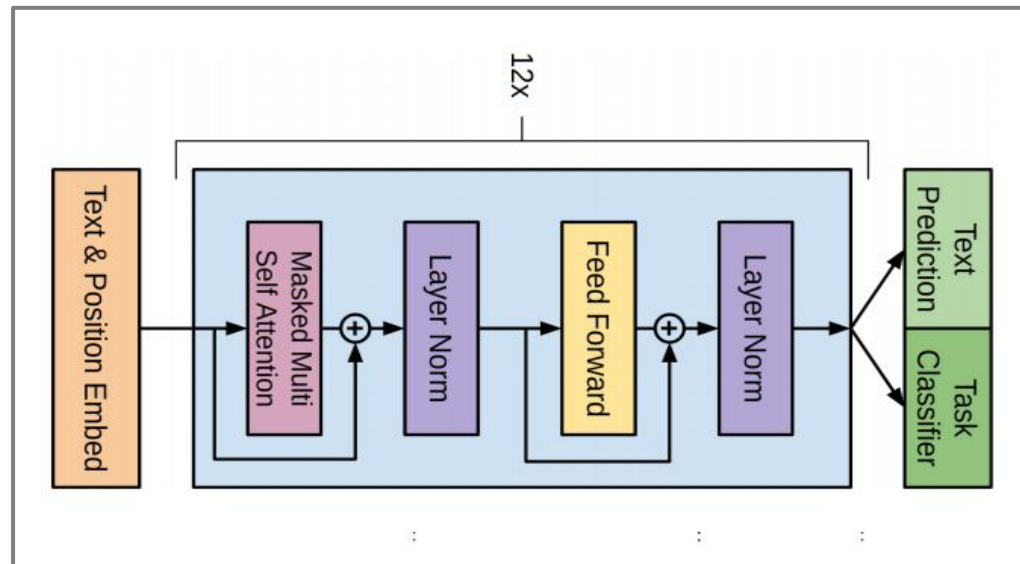
LSTM 모형 (Long Short-Term Memory)

RNN의 단점을 보완한 모델이며, 은닉층의 메모리 셀에 입력, 망각, 출력 게이트를 추가하여 불필요한 기억을 지우고 기억해야 할 것들을 저장하여 긴 시퀀스의 입력을 처리 가능



GPT2

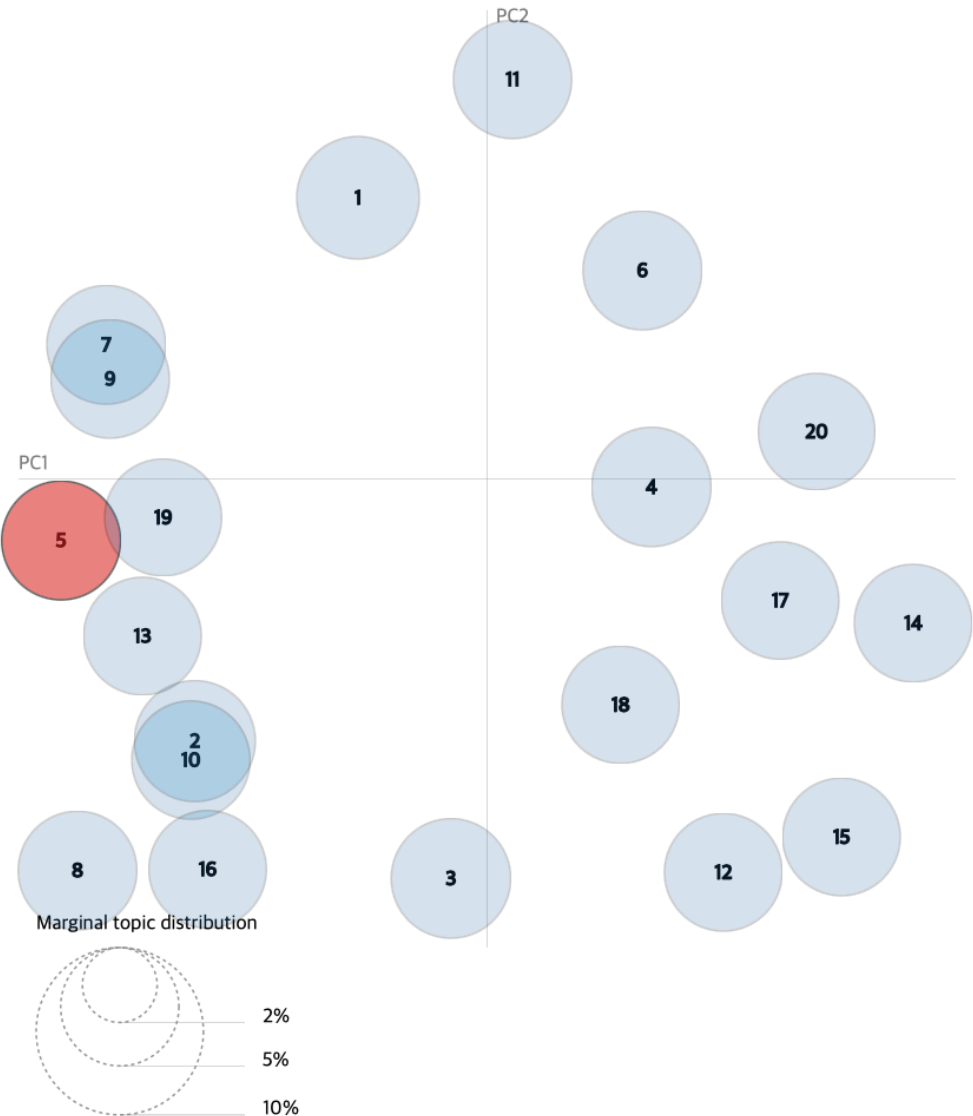
transformer decoder를 활용한 pretrained LM 이다.
pretrained LM이란 레이블이 없는 데이터를 비지도 학습 방법으로 학습해서 모델이 언어를 이해 할 수 있도록 한 후
특정 Task에 적용해서 좋은 성능을 내는 방법



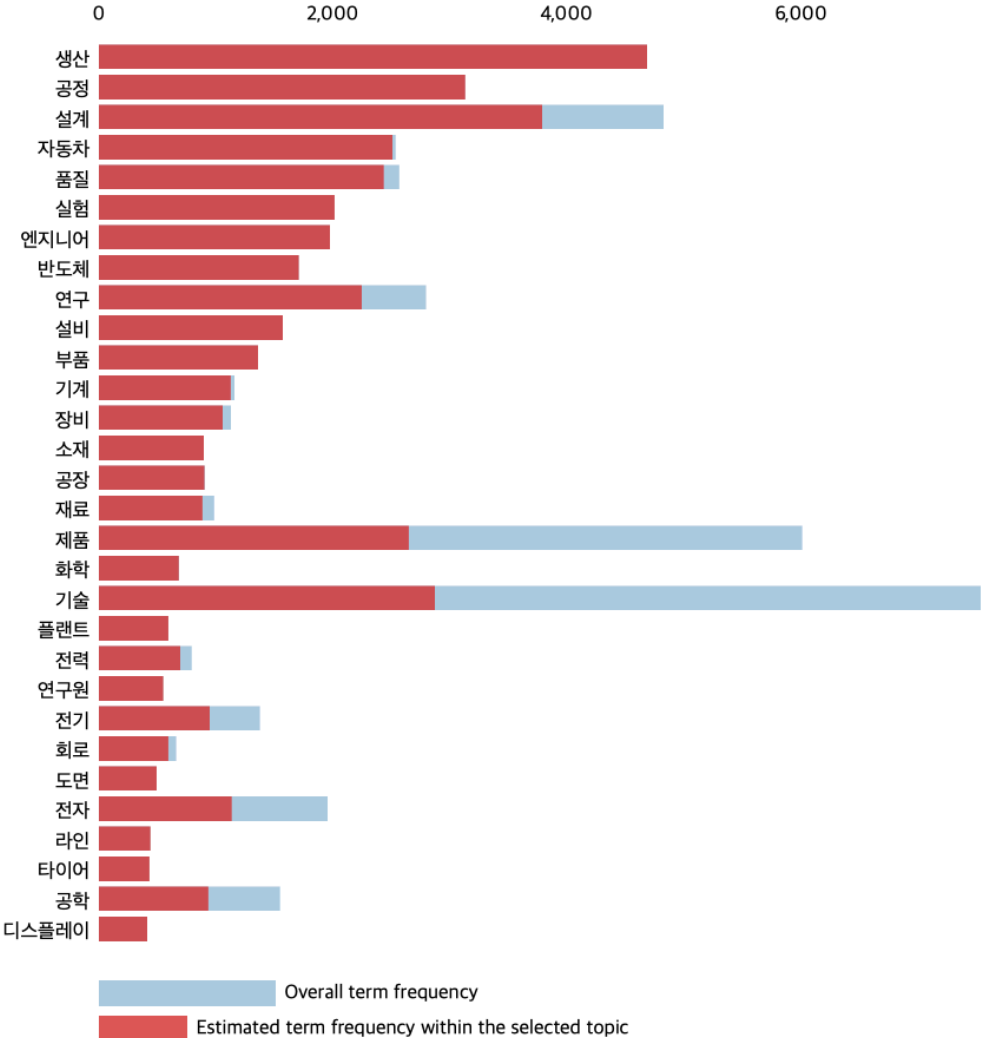
단어 및 소제목

5번 토픽
기계, 전기전자

Intertopic Distance Map (via multidimensional scaling)



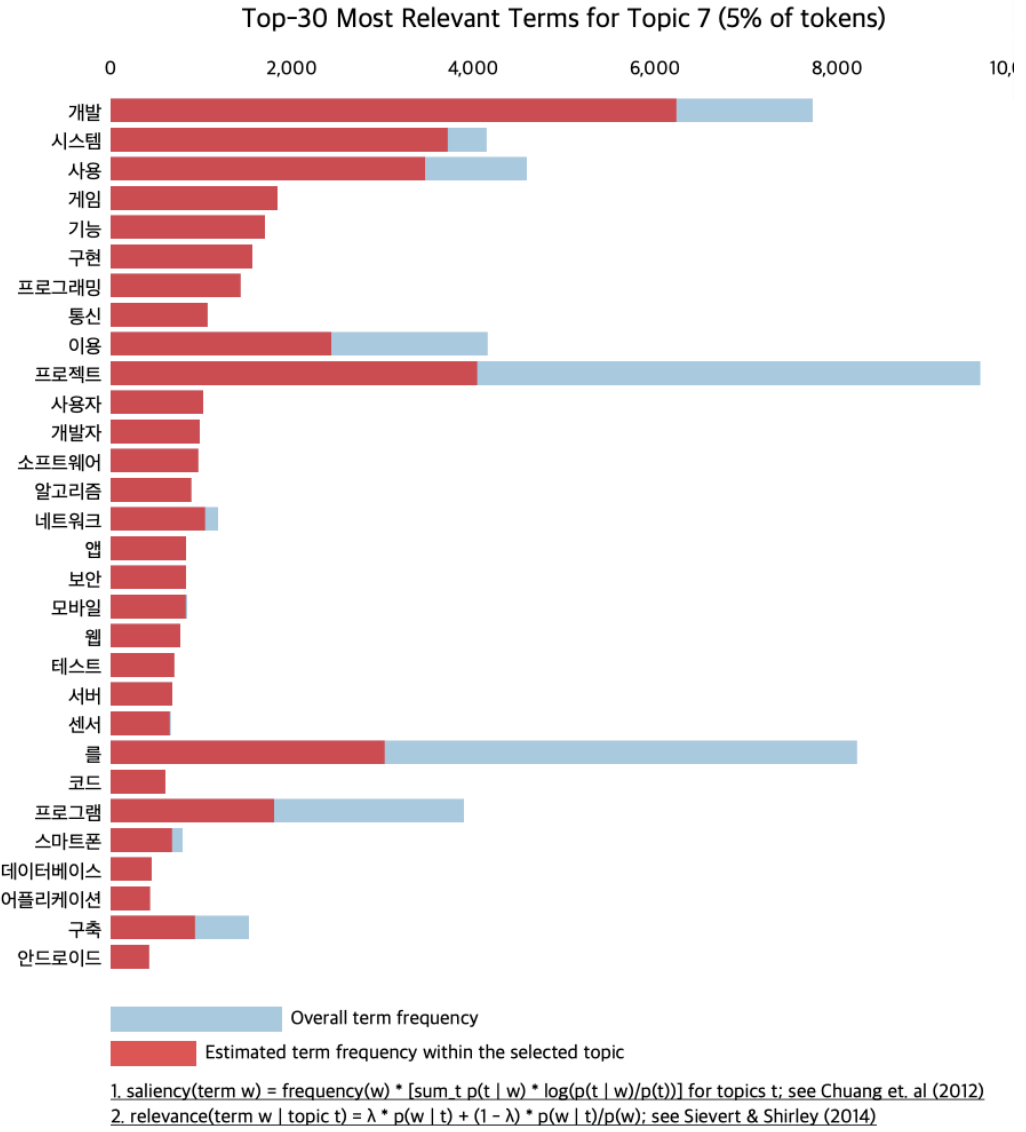
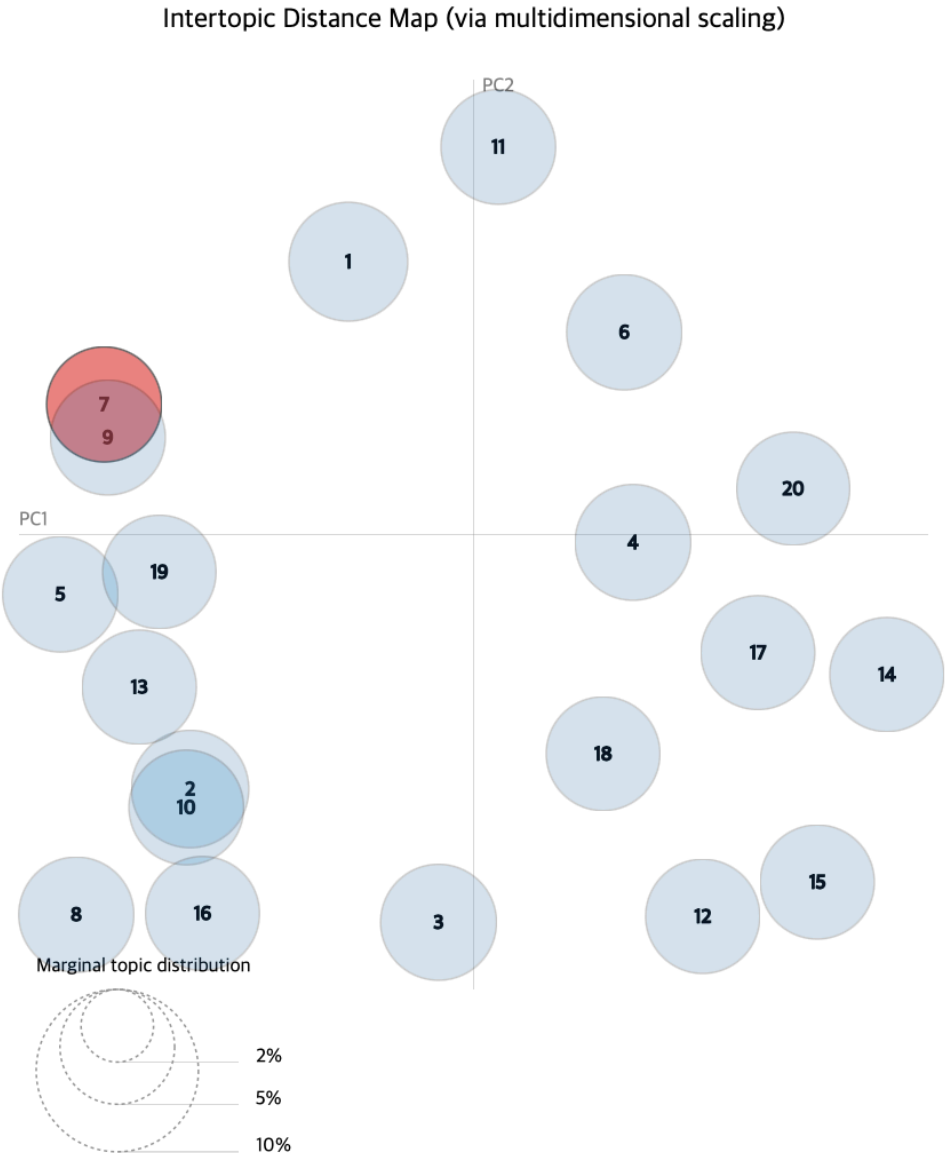
Top-30 Most Relevant Terms for Topic 5 (5% of tokens)



1. $saliency(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)
2. $relevance(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

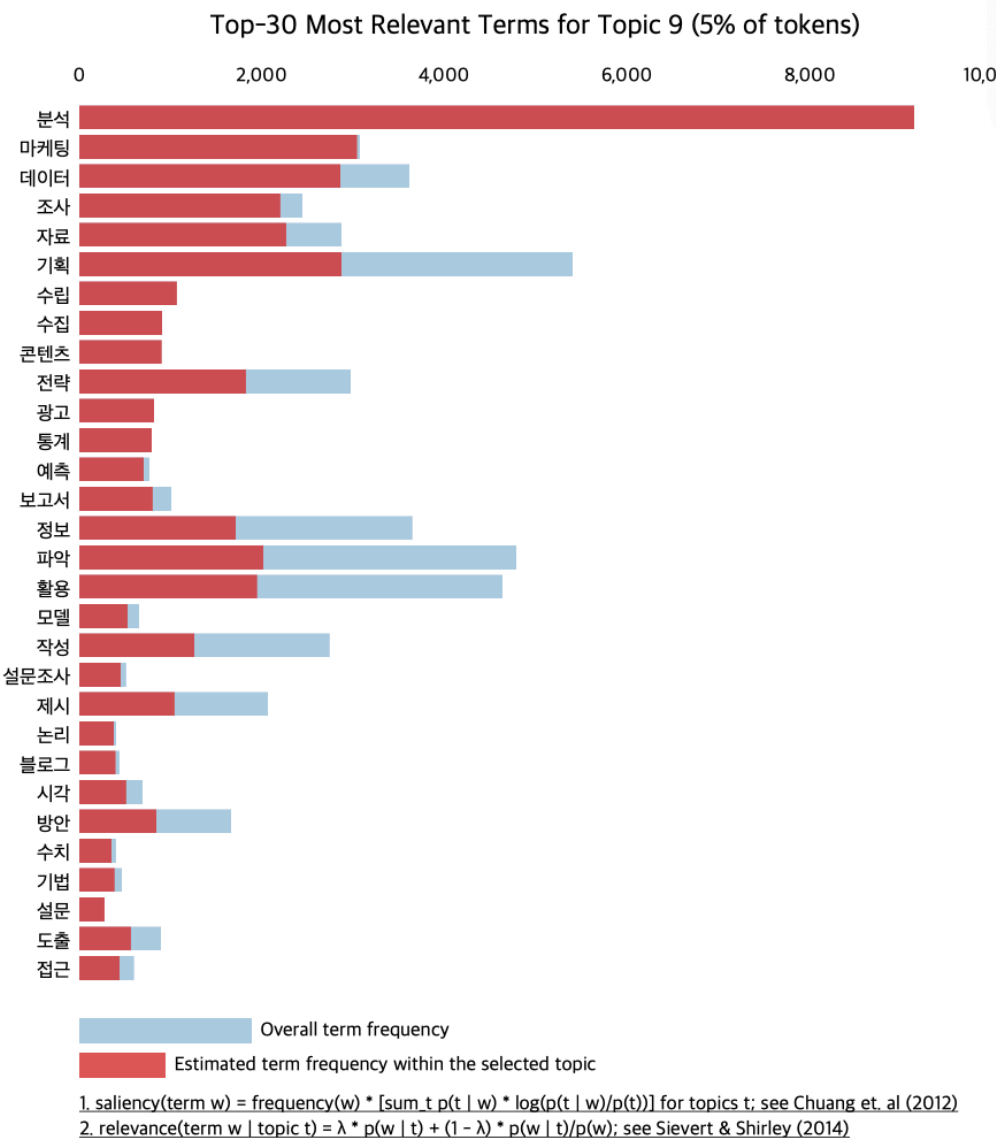
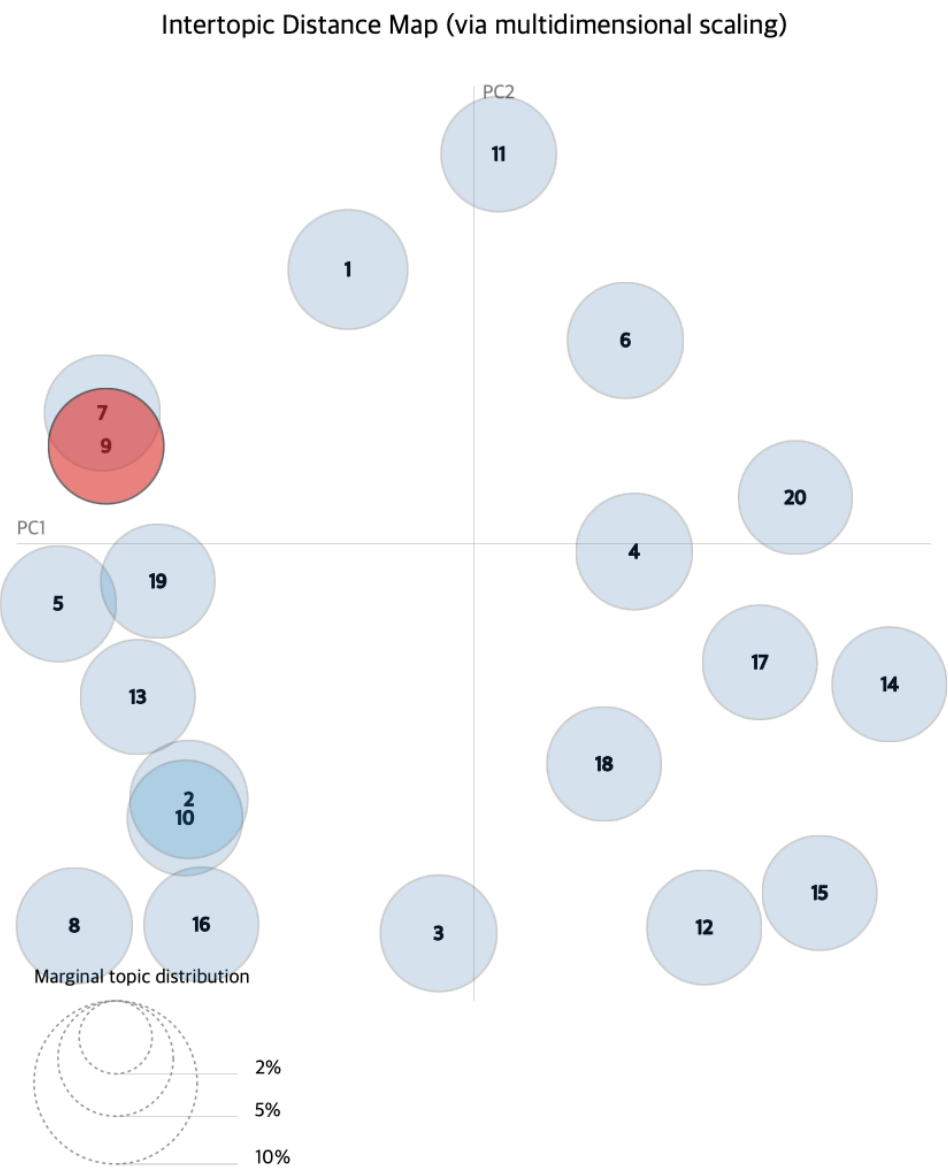
단어 및 소제목

7번 토픽
IT, 개발



단어 및 소제목

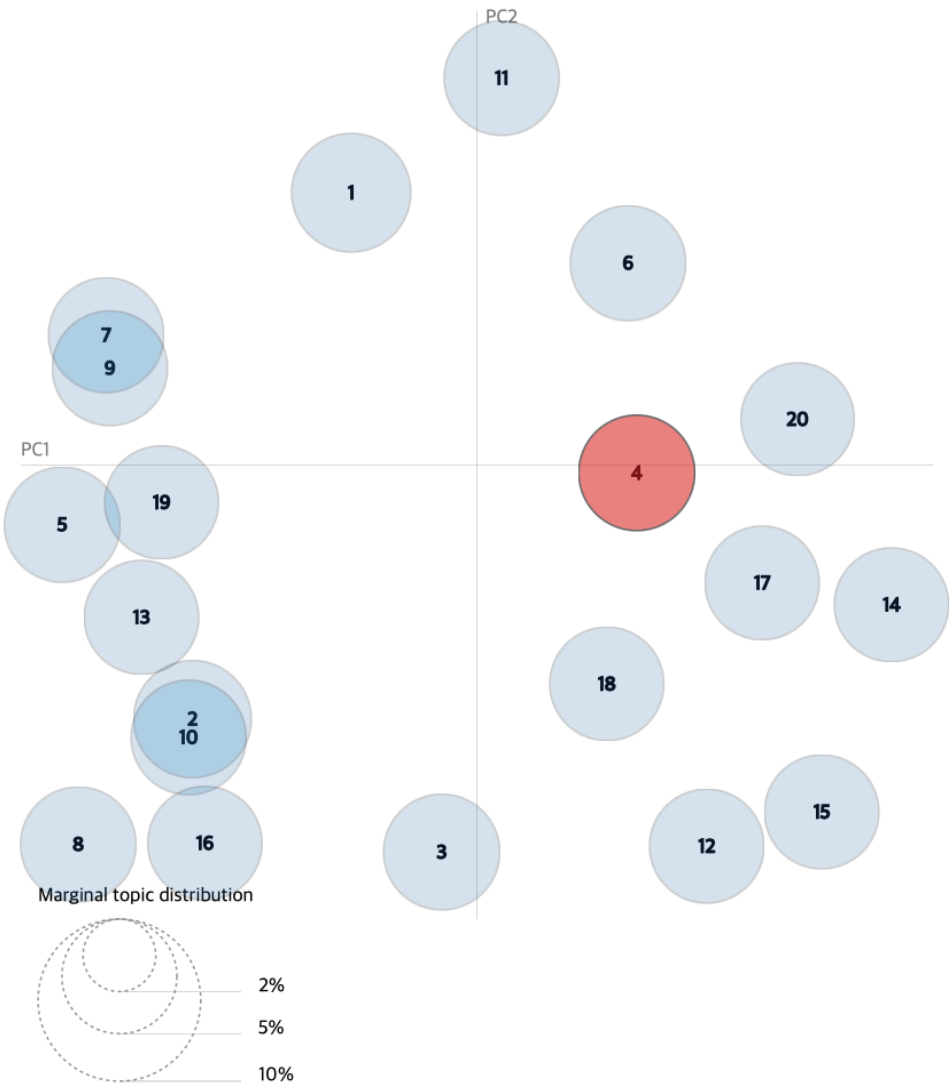
9번 토픽
영업, 마케팅



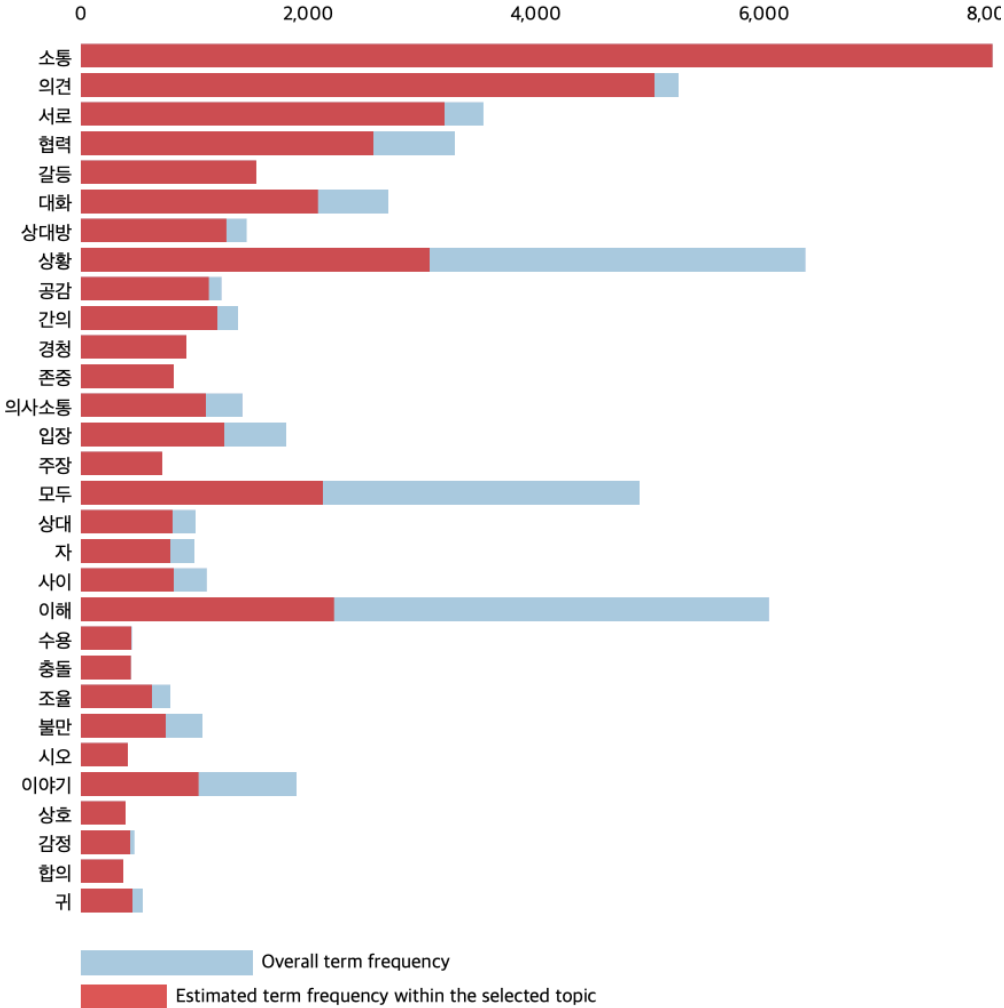
단어 및 소제목

4번 토픽
문제해결능력

Intertopic Distance Map (via multidimensional scaling)



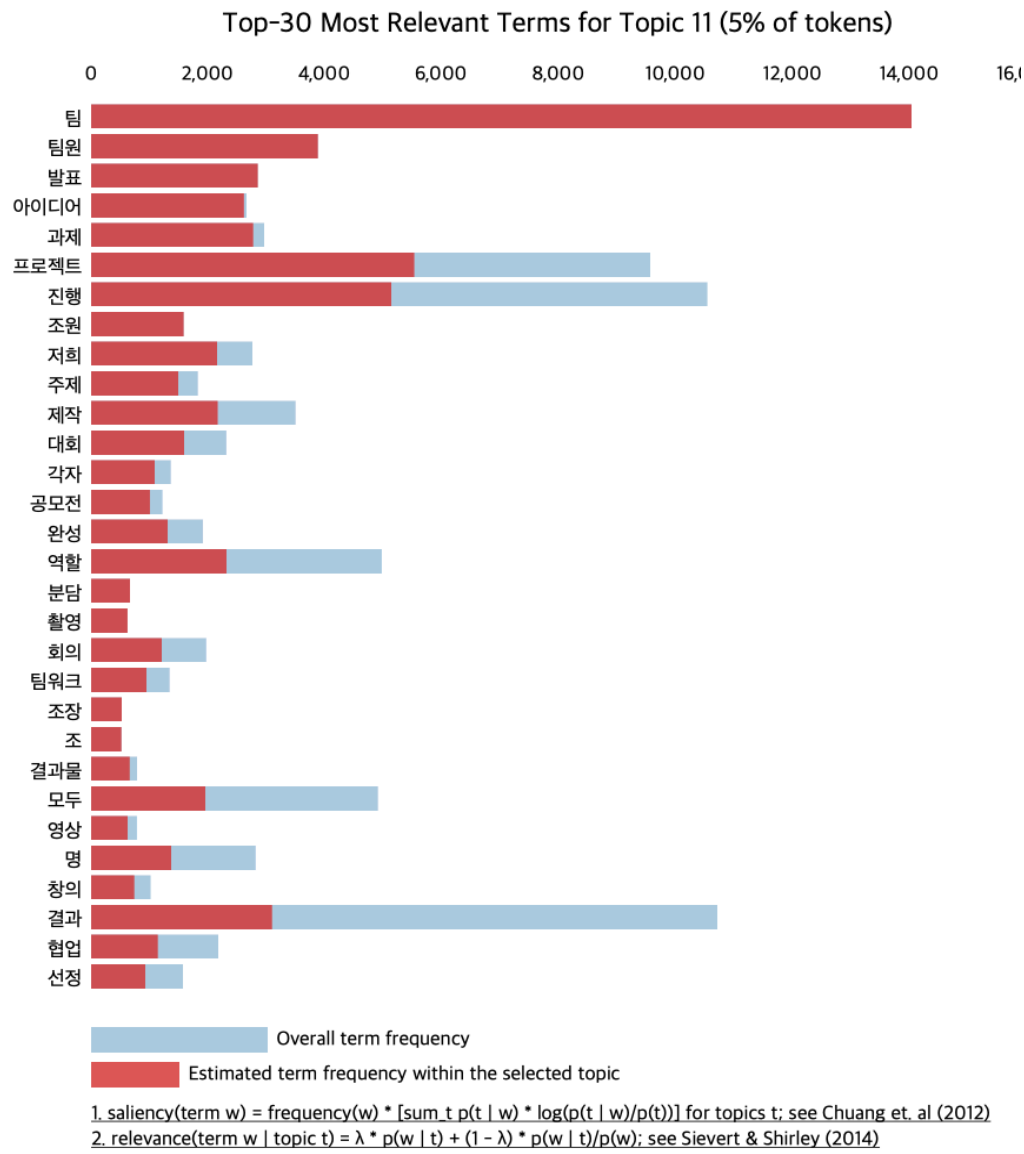
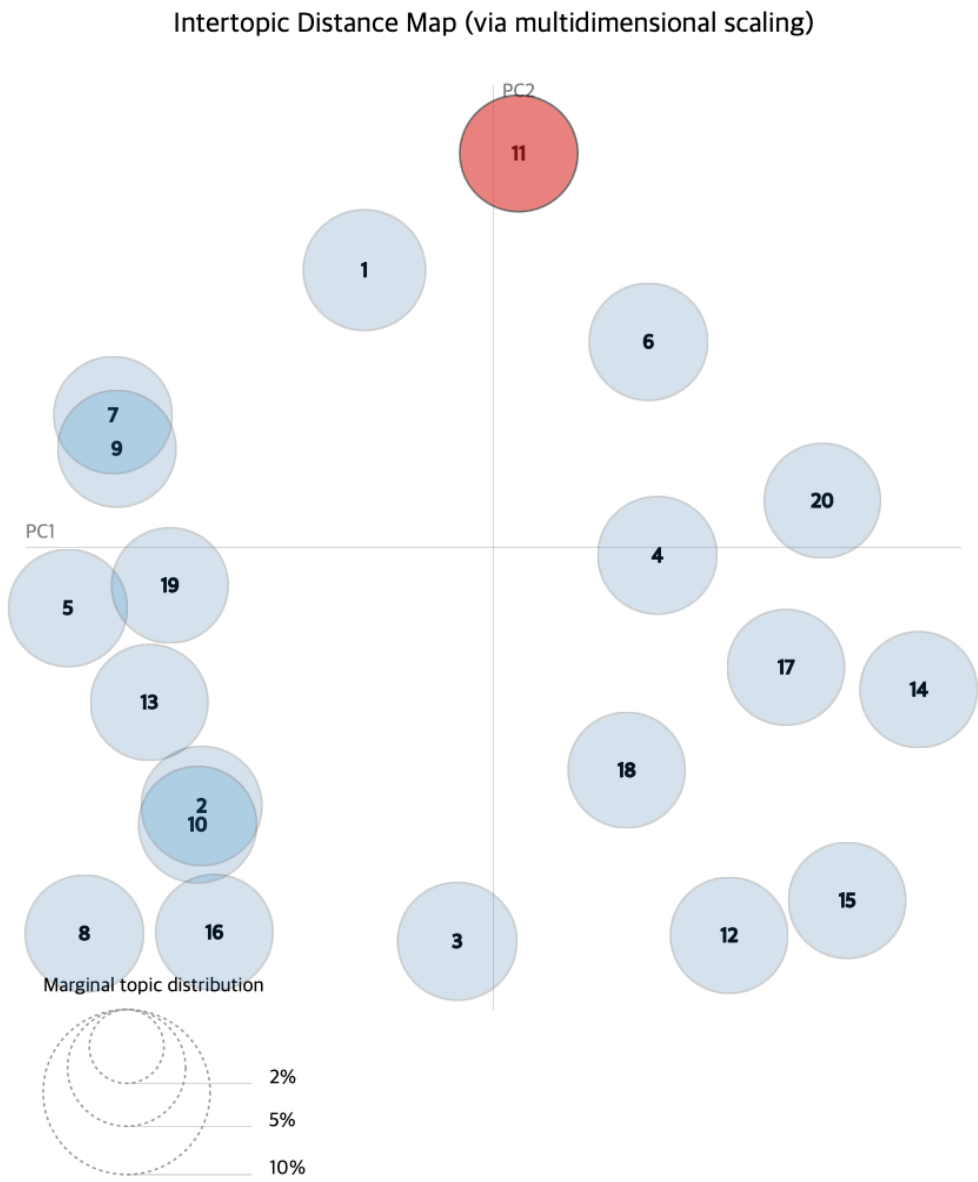
Top-30 Most Relevant Terms for Topic 4 (5.1% of tokens)



1. saliency(term w) = frequency(w) * [sum t p(t | w) * log(p(t | w)/p(t))]] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

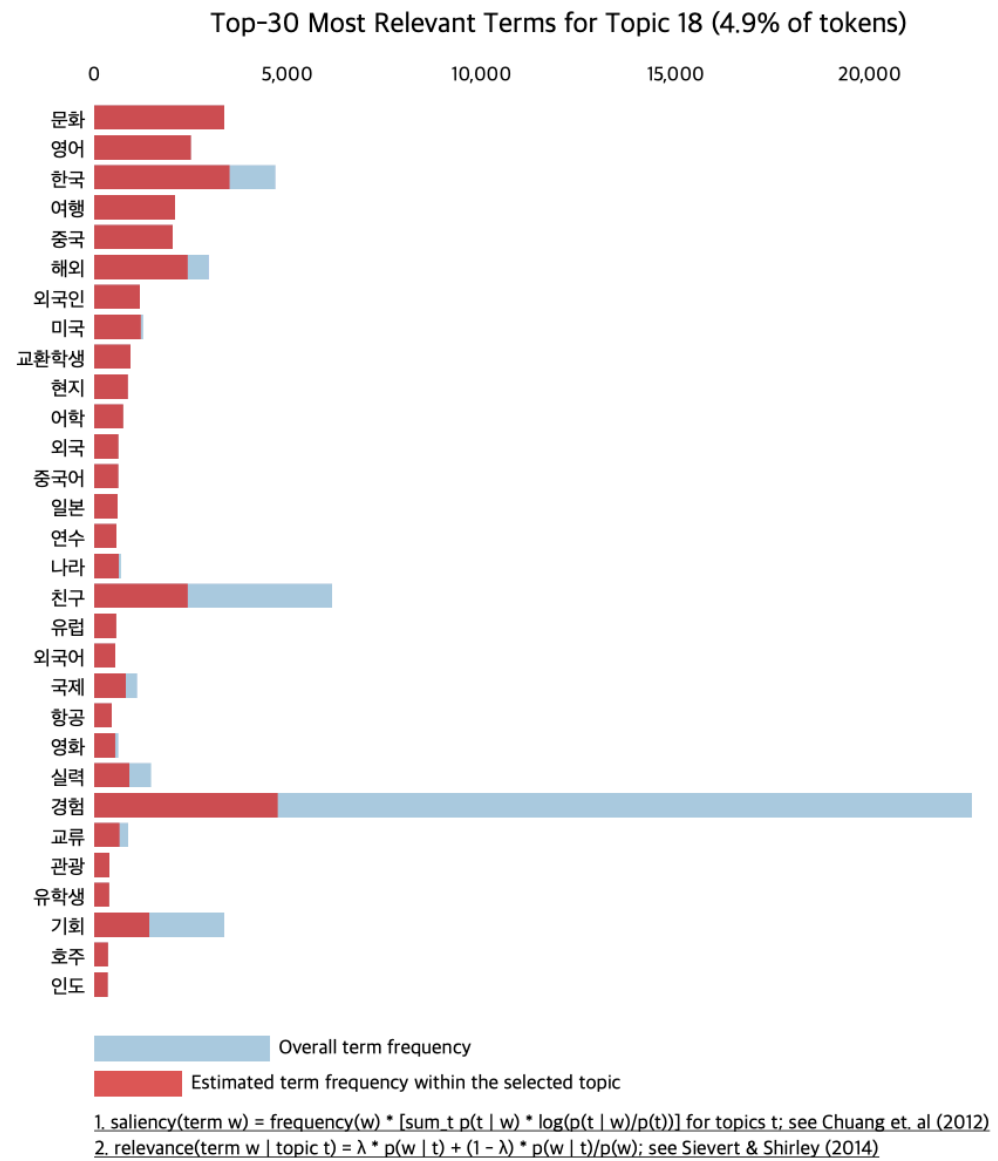
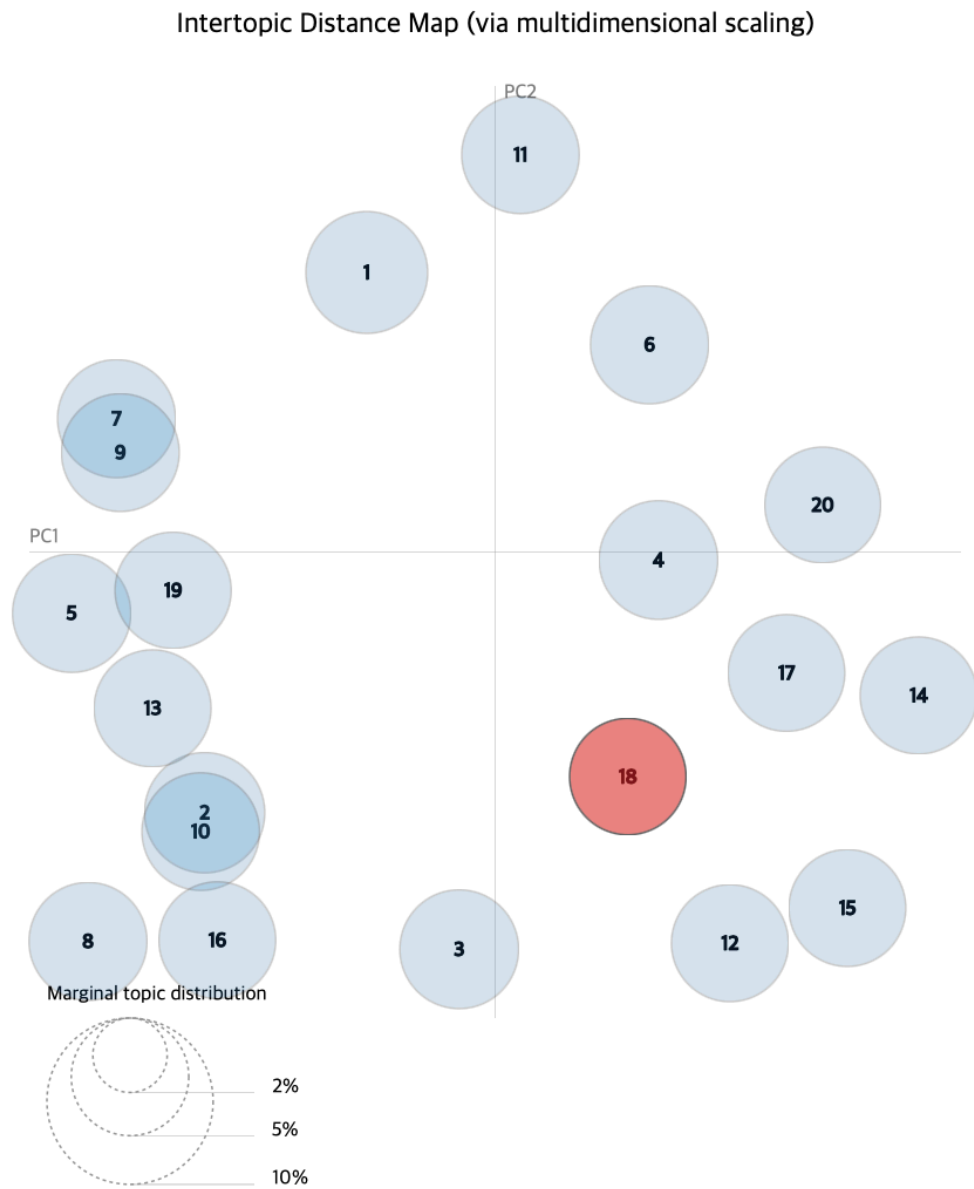
단어 및 소제목

11번 토픽
대외활동



단어 및 소제목

18번 토픽
글로벌 역량



소제목 작성

Keyword : 봉사

RNN

“봉사 프로젝트 협력을 통해 만드는 성장”

LSTM

“봉사 경험을 통해 얻은 인재”

KoGPT2

“봉사와 열정 에너지”
“성실함과 깐깐함”
“고객 니즈 파악은 나의 연결고리”

소제목 작성

Keyword : 개발

RNN

“개발 경험 산업의 삶의 인재”

LSTM

“개발 동아리 신입 향상”

KoGPT2

“개발 인생에서의 도전 !”
“새로운 것에 대한 관심”
“협력적 마인드와 팀워크를 이루는 영업인”

소제목 작성

Keyword : 데이터

RNN

“데이터 분석 중 개발이 있어야 한다”

LSTM

“데이터 분석 경험을 바탕으로 한 능력”

KoGPT2

“데이터의 미래라는 가치관을 가진 인재가 되겠습니다”
“배려, 미래를 알다. 사람을 알고, 알다”

소제목 작성

Keyword : 직무

RNN

“직무 역량 강화를 위한 노력”

LSTM

“직무 관련 경험을 갖춘 성장”

KoGPT2

“직무 지원동기 최고의 가치는 정직함 새로운 가치를 만드는 사람”
“고객의 니즈 파악하는 사람 정직함을 갖추었습니다”
“경험을 통해 준비된 인재”

소제목 작성

Keyword : 노력

RNN

“노력 혁신은 작은 곳에서 출발한다.”

LSTM

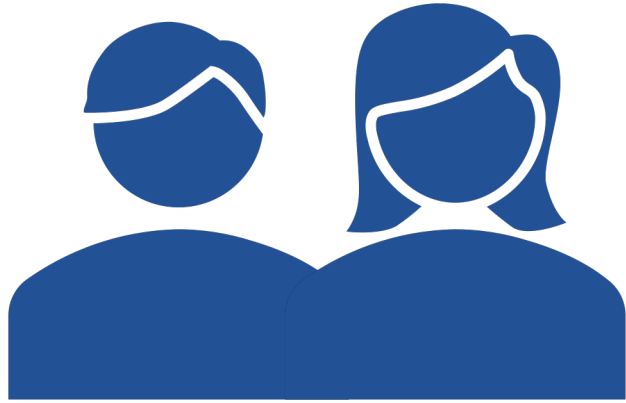
“노력 많은 엔지니어 운영”

KoGPT2

“노력을 통한 팀워크의 중요성”

“사람을 알고, 일을 알고, 함께라는 의미를 실천하는 인재”

“사람의 마음을 사로잡은 아이디어 बैं크”



취업준비생 측면

1. 특정 분야 및 기업에 관한 키워드 제공
2. 막힌 단어 추천 제공
3. 자기소개서 중요 문장 제공
4. 소제목 작성 제공



기업 측면

1. 채용 의사 결정에 대한 도움
2. 자기소개서 내용 파악 시간 단축
3. 취업준비생에게 피드백을 제공하는 서비스 창출

한계점

1. 데이터가 너무 커서 학습 시간이 너무 많이 소요되어서 학습을 많이 진행하지 못해 결과가 매끄럽지 않은 점
 2. 소제목뿐만 아니라 내용에 관해서도 생성을 해보고 싶었지만 컴퓨터의 성능이 따라오지 못하여 학습 중단이 된 점
 3. GPT3가 최신 모델이지만 관련 자료가 부족하여 사용하지 못한 점
- 추후에는 더 많은 데이터로 더 많은 학습과 최신 모델로 학습을 해보고 싶다.
 - 자기소개서 작성에 도움을 주는 웹사이트를 제작하고 싶다.



감사합니다

