
GUARDIANS OF DATA

- +
 - BERT를 이용한
○ NEWS 요약 및 추천 시스템

오수연 박소연 박병준 김학진 제태성
With Mentor 이성희

CONTENTS

+
•
○

01 주제소개

02 팀원소개

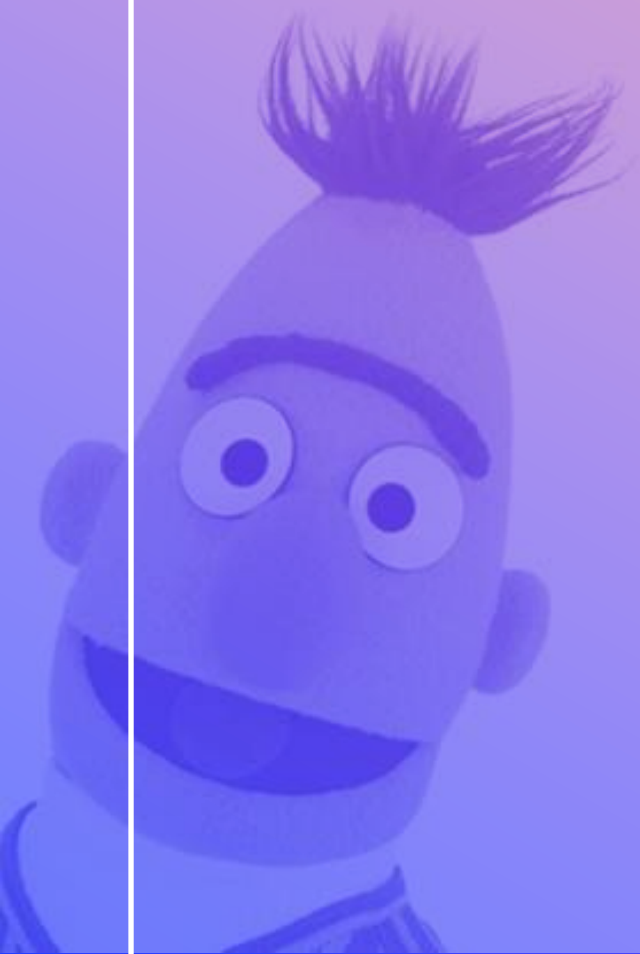
03 데이터셋

04 아키텍처

05 알고리즘

06 성능평가

07 기대효과



Google
BERT



INTRODUCTION

무엇을 할 것인가?

Background of Topic Selection



시스템 목표

최신 뉴스와 가장 연관성 있는 기존 뉴스 추천

추출된 요약문과 기존 뉴스와의 연관도 계산

최신 뉴스 추출 요약문 생성

최신 뉴스 검색

TEAM



이성희
Mentor

Team Leader



오수연

Embedded Vector 데이터 셋 구축
Vector 유사도 계산 모델 개발



박소연

Data Pre-processing
Deep Learning Modeling



박병준

Data Pre-processing
Deep Learning Modeling



김학진

Database 구축
Front-end 구축



제태성

Embedded Vector 데이터 셋 구축
Presentation 및 산출물 작성



DATA SET

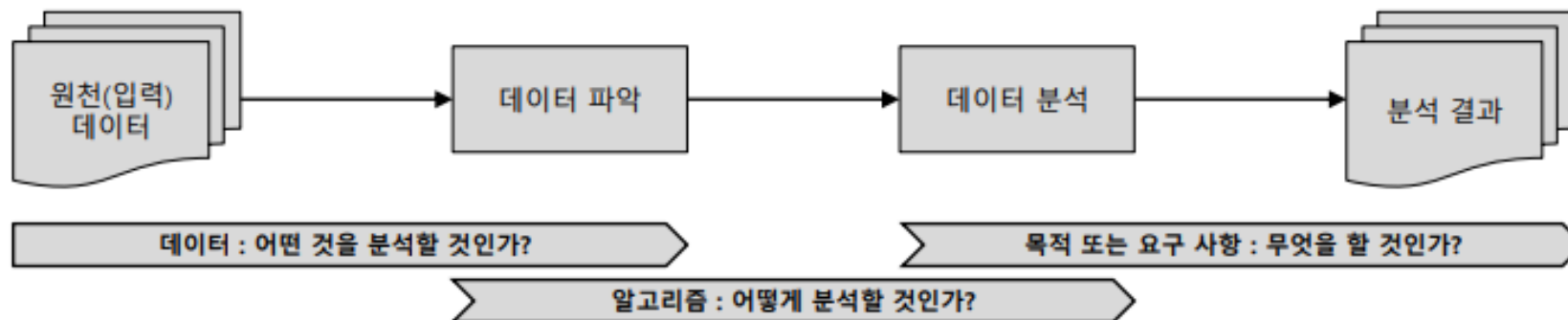


무엇을 분석할 것인가?
쓸모 있는 데이터를 찾아라!

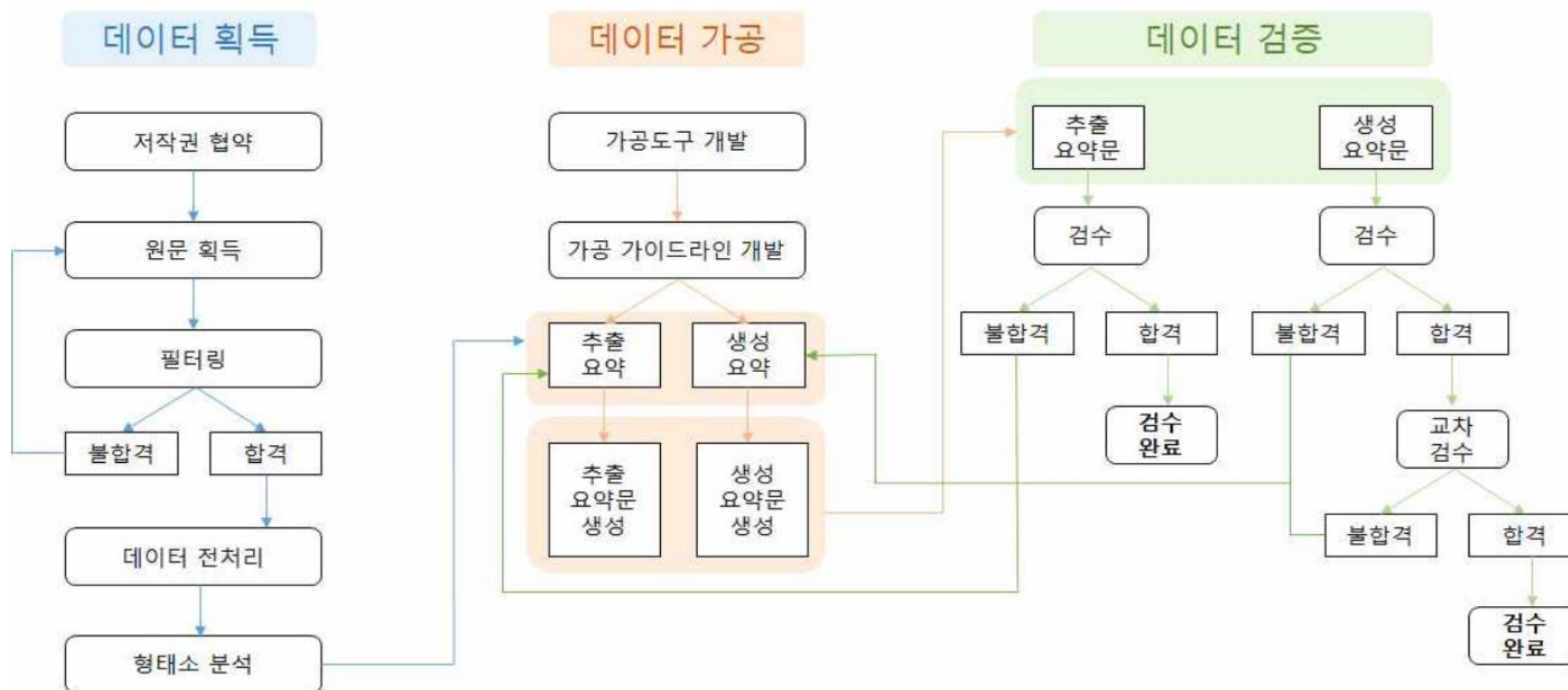
DATA

- 데이터 분석을 위한 3가지 정의

- 무엇을 할 것인가?
- 무엇을 분석할 것인가?
- 어떻게 분석할 것인가?



DATA 구축 개요



ROW DATA

원문 데이터
신문기사 30만 건

추출 요약
3개 문장

+

#문서요약

#추출요약

#생성요약

문서요약 텍스트

분야

한국어

유형

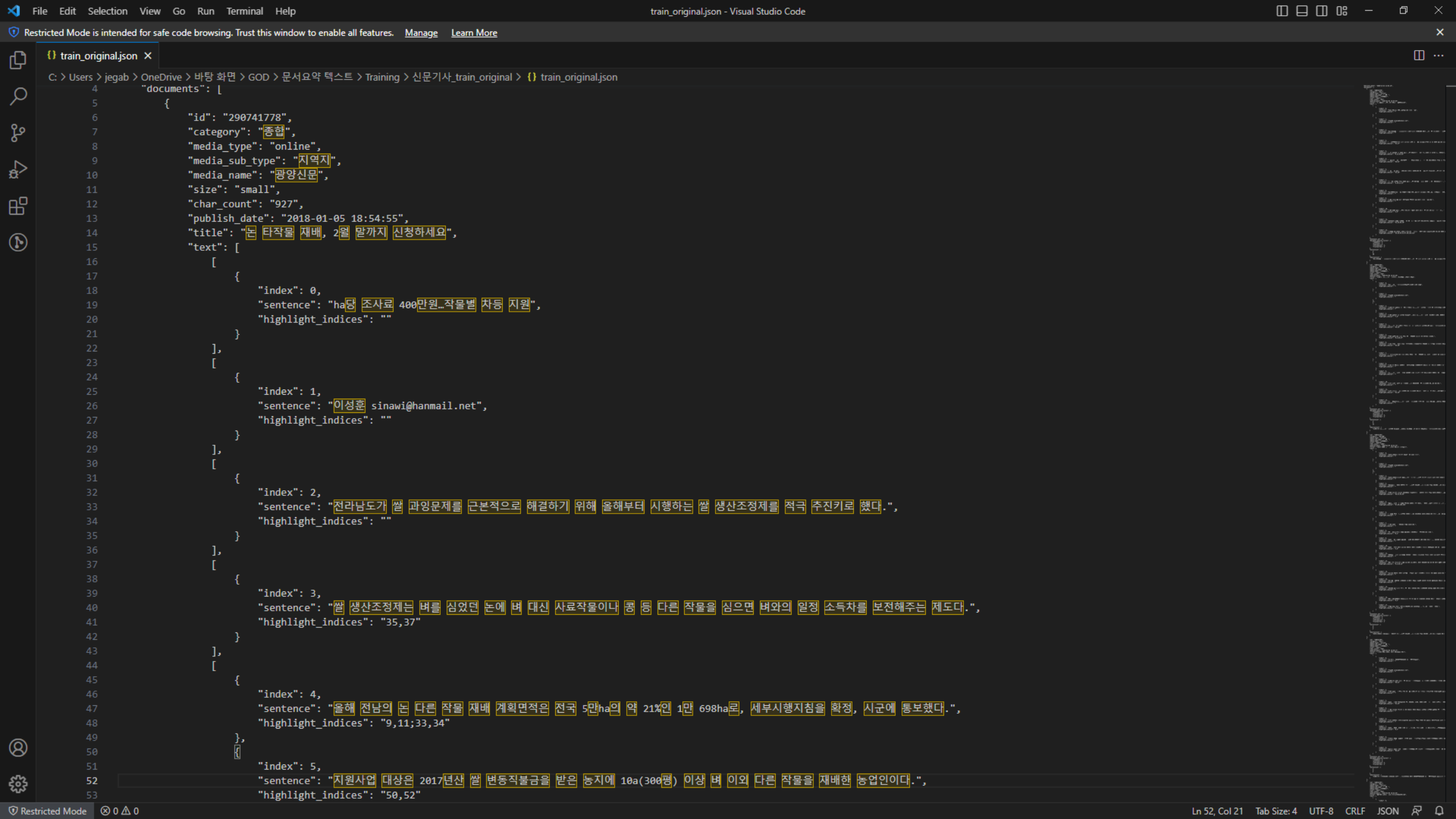
텍스트

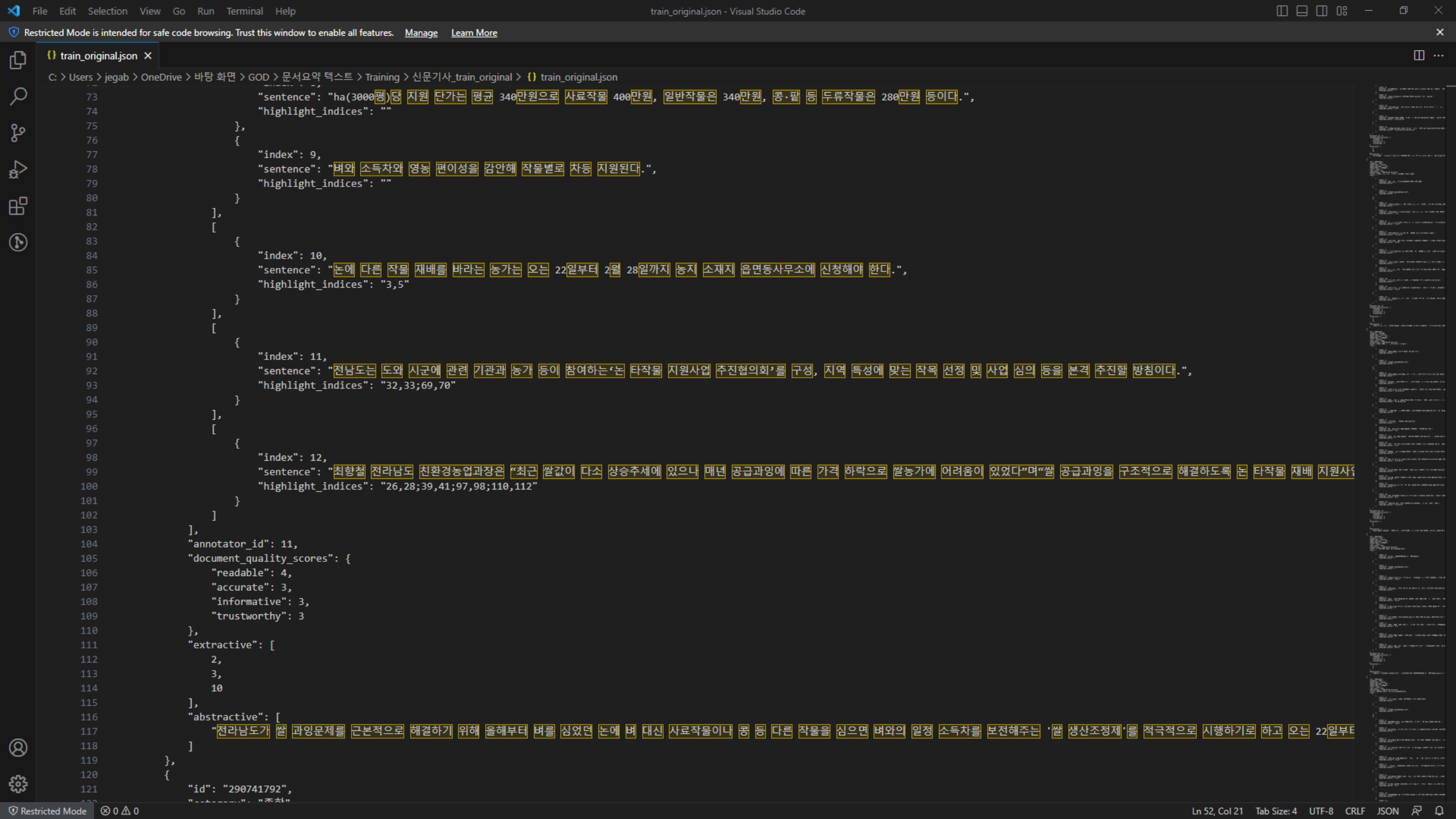
갱신년월 : 2023-05 구축년도 : 2020

다운로드



샘플 데이터





DATA PRE-PROCESSING

쓸 수 있는 데이터로 변환하라!

dataset_final_eco.csv X dataset_final.csv

dataset_final_eco.csv

1 sentence,target,category
2 중국 투자의향기업 직접 방문·투자협약 체결 협의,0,경제
3 이성훈 sinawi@hanmail.net,0,경제
4 여수광양항만공사(사장 방희석)는 14일부터 20일까지 광양항 배후단지 투자 유치에 CEO가 직접 참여하는 투자유치 활동을 중국 중남부지역에서 펼친다.,1,경제
5 광양만권경제자유구역청과 합동으로 진행되는 이번 투자유치활동은 방희석 사장이 직접 중국 현지의 투자의향기업을 방문해 광양항 및 배후단지의 장점 등을 소개하고 투자협약(MOU)을 체결하는 방식으로 진행된다.,1,경제
6 방 사장은 먼저 광양항 서측배후단지 푸드존 투자유치를 위해 중국 유남성 소재 커피 원재료 공급업체인 운남허메이저치(주)를 방문해 광양항 배후단지 투자협약을 체결할 예정이다.,0,경제
7 이어 한중일 콜드체인 거점 구축을 위해 중국 안후이성 소재 냉동냉장창고 업체인 바이란식품(주)을 방문해 투자협약을 맺기로 했다.,0,경제
8 중국의 다른 입주의향기업들도 방문해 투자유치 활동을 진행할 계획이다.,0,경제
9 "이번 투자유치 활동은 지난해 9월 경제청과 공사가 합동으로 중국지역 투자유치활동을 통해 투자의향기업을 선정했던 후속 조치로, 이번 방문을 통해 광양항 배후단지에 실질적인 투자가 이뤄질 수 있도록 하기 위한 것이다.",1,경제
10 "방희석 사장은이번 중국 방문은 공사, 경제청이 합동으로 광양항 배후단지에 대한 투자유치를 이끌기 위한 것이라며광양항 서측배후단지 푸드존에 중국기업이 입주하는 실질적인 투자유치 성과가 있을 것으로 기대된다고 배경을 설명했다.",0,경제
11 방 사장은광양항 배후단지를 한·중·일 콜드체인 거점 및 식품산업 클러스터로 발전시키는 충분한 계기가 될 것이라며앞으로도 광양항 배후단지 활성화를 위해 CEO가 직접 투자유치 활동에 나서는 등 가시적인 효과를 낼 수 있도록 최선을 다하겠다고
12 이성훈 sinawi@hanmail.net,0,경제
13 포스코 광양제철소(소장 김학동)가 자동차 강판 생산공장인 제3냉연공장을 최신 설비로 재탄생시키며 자동차 강판 생산 전문 제철소로의 경쟁력을 더욱 강화했다.,0,경제
14 광양제철소는 지난 16일 3냉연공장 합리화 준공식을 열었다.,1,경제
15 "광양제철소가 최신설비로 바꾼 까닭은 최근 자동차 업계에서 각종 환경과 안전 규제 등으로 인해 가볍고 단단한 초고강도강판의 수요가 증가하고, 중국 시장을 중심으로 지속적으로 늘어나는 자동차 판매 수요에 대응하기 위함이다.",1,경제
16 "지난 10월부터 약 3개월 간 진행된 공사는 철을 잡아당겼을 때 끊어지는 인장강도가 1기가파스칼 이상인 기가급 초고강도 강판 생산과, 자동차 강판 중 높은 품질기술력을 요구하는 자동차외판 생산 중대에 초점을 두고 진행됐다.",0,경제
17 "기가급 초고강도강판은 외부 충격을 튼튼하게 막아줄 자동차내부 뼈대인 프레임에 쓰이며, 자동차외판은 차량 문이나 보닛 등 외부에 노출되는 면에 쓰인다.",0,경제
18 "1992년 첫 가동을 시작한 3냉연공장은 지난해까지 4900만톤의 냉연제품을 생산해왔으나, 설비 성능상 기가급 초고강도강판은 생산하지 못했다.",0,경제
19 하지만 치열해지는 자동차 강판 시장에서 제철소의 종합 경쟁력을 높이기 위해 900여억원을 투자해 고강도와 고품질의 제품을 생산할 수 있도록 설비 성능을 대폭 끌어올렸다.,0,경제
20 "이로써 4개 냉연공장 중 1,2냉연공장에 이어 3공장도 기가급 초고강도강판을 생산할 수 있는 기반을 마련하게 됐다.",0,경제
21 "3냉연공장은 올해 연 3만톤 가량의 기가급 초고강도강판을 생산할 계획이며, 2020년 연간 10만톤 생산을 목표로 하고 있다.",0,경제
22 아울러 업그레이드된 설비를 통해 차외판 생산량은 작년 5만톤 대비 올해 50% 이상 늘릴 예정이다.,0,경제
23 이를 통해 광양제철소는 자동차 강판 등 포스코 주력 제품인고부가가치(WP) 제품 생산 확대에 안정적으로 대응할 수 있을 것으로 보고 있다.,1,경제
24 김학동 제철소장은이번 공사를 통해 3냉연공장은 WP제품 생산 능력을 강화하게 돼 포스코의 미래 경쟁력을 한층 더 높게 됐다고 평가했다.,0,경제
25 김 소장은3냉연공장에 몸을 담고 있는 모든 포스코 패밀러 구성원들이 세계 최고를 지향하는 마음으로 끊임없이 노력해 포스코 미래 50년을 위한 중추적인 역할을 해주길 기대한다며공사가 성공적으로 마무리되도록 노력해준 모든 관계자분들에게 깊은
26 광양제철소는 3냉연공장 합리화 사업 준공을 바탕으로 WP제품 생산능력 확대 및 품질 향상을 위한 설비 고도화를 지속 추진할 계획이다.,0,경제
27 이와 함께 4차 산업 기술을 활용한 스마트 팩토리 구현으로 Strong & Smart 제철소실현을 점차 앞당길 예정이다.,0,경제
28 이성훈 sinawi@hanmail.net,0,경제
29 "여수광양항만공사(사장 방희석, 이하 공사)는 2018년도 승진 및 전보 인사를 단행했다고 지난 26일 밝혔다.",1,경제
30 "공사는 광양항 활성화 및 물동량 창출을 위해항만운영팀, 글로벌마케팅팀, 물류단지팀의 인력을 보강하고자 인사를 실시하게 됐다고 설명했다.",1,경제
31 특히 정부 정책의 적극적인 이행과 일자리 창출 등 사회적 가치 실현을 위해가치경영팀을 신설한 것도 이번 인사의 요인이라고 덧붙였다.,1,경제
32 *2018년도 승진·전보 인사,0,경제
33 <승진> ◇ 1급(실·팀장) △항만운영팀장 김한준 ◇ 2급(부장) △항만개발팀장 고영한 △재무회계팀 박정철 △물류단지팀 장방식 ◇ 3급(차장) △가치경영팀 박상우 △인재육성팀 박진수 △기획조정실 양현모 △항만시설팀 고훈정 ◇ 4급(과장) △재무
34 <전보> ◇ 1급(실·팀장) △재무회계팀장 최연철 △항만개발팀 기술자문위원 권영원 ◇ 2급(부장) △기획조정실장 선정덕 △가치경영팀장 정기철 △글로벌마케팅팀장 백정원 △미래사업팀 조성래 △항만운영팀 천철웅 ◇ 3급(차장) △인재육성팀 윤승기
35 지역 특화작목>,0,경제
36 "특새소득작목 발굴, 미래농업 준비",0,경제
37 이성훈 sinawi@hanmail.net,0,경제
38 광양시는 경쟁력을 갖춘 품으로온 농어촌 건설을 위해 원예작물 신기술보급사업을 통한 미래 신성장동력 육성에 나설 방침이다.,1,경제
39 "시는 그동안 농산물 과잉 생산과 시장 개방, 기후 변화, 농업 경영비 증가 등 급변하는 농업환경에 선제적으로 대응하기 위해 특화작목과 특새소득작목 발굴·육성 등 신기술보급사업을 추진해 왔다.",0,경제
40 "올해는 기후변화 대응 아열대 소득과수 시범 도입과 생강, 취나물, 수박, 알스트로메리아 등 지역 특화작목을 비롯한 미래농업 분야에 집중투자할 계획이다.",1,경제
41 "시는 올해 아열대소득과수 육성을 위해 생산과 유통, 가공 등 6차산업화를 추진할 계획이다.",0,경제
42 "농업기술센터 시험포장에 1000㎡규모의 열대과수관을 설치해 애플망고 등 12품목을 시험재배하고, 품목 발굴과 재배기술 체계 확립에 나설 계획이다.",0,경제
43 광양생강을 지역 특화작목을 넘어 전국 최고의 특화작목으로 육성하는 계획도 추진한다.,0,경제
44 시는 2016년 전라남도농업기술원 공모사업인1시간 1특화작목 육성 사업에 3개년동안 사업비 10억 원을 확보하고 2019년까지 집중적으로 투자할 계획이다.,0,경제
45 "지난해까지는 생산 위주의 사업으로 고품질 생강 재배단지 조성, 생강 비가림 시설재배 실증시험에 집중적으로 투자한 결과 2016년 16.3ha에서 2017년 18ha로 증가하는 성과를 거뒀다.",0,경제
46 "브랜드 규격 박스를 통한 농협 계통출하로 상품의 청결성과 신뢰도를 높였으며, 지난해 11월 8일 10kg 상품 기준 광양 생강 최고 도매가격은 50,000원, 광양 평균 도매가격은 4만3600원으로 전국 평균 3만2300원보다 높게 나타났다.",0,경제
47 "취나물, 수박, 알스트로메리아 등 지역 특화작목들도 꾸준히 육성한다.",0,경제
48 "시는 고품질 취나물 생산을 위해 비가림하우스 시설을 0.3ha로 확대할 계획이며, 시설수박 단지 82ha에는 품질향상과 안전한 농산물 생산을 위해 수정별을 활용한 자연수정기술을 지원할 계획이다.",0,경제
49 "또 알스트로메리아 등 화훼류에는 생산비 절감 기술이 투입되며, 취나물과 화훼류에는 규격 박스를 지원해 상품성을 높여나갈 방침이다.",0,경제
50 "이밖에 대표적인 효자 작목인 매실과 애호박 집중 육성, 특새소득작물 발굴을 위해 딸기 0.4ha와 왕도란 0.7ha 재배 등 기존 농업의 새로운 활로를 찾아나갈 계획이다.",1,경제
51 허명구 기술보급과장은시는 광양농업의 미래에 선제적으로 대응하기 위해 원예작물 신기술보급 분야를 점차적으로 확대하고 있다며올해는 아열대과수와 생강 특화작목 육성을 적극 추진해 시의 대표 농특산물로 자리매김할 수 있도록 최선을 다하겠다고

TABLE

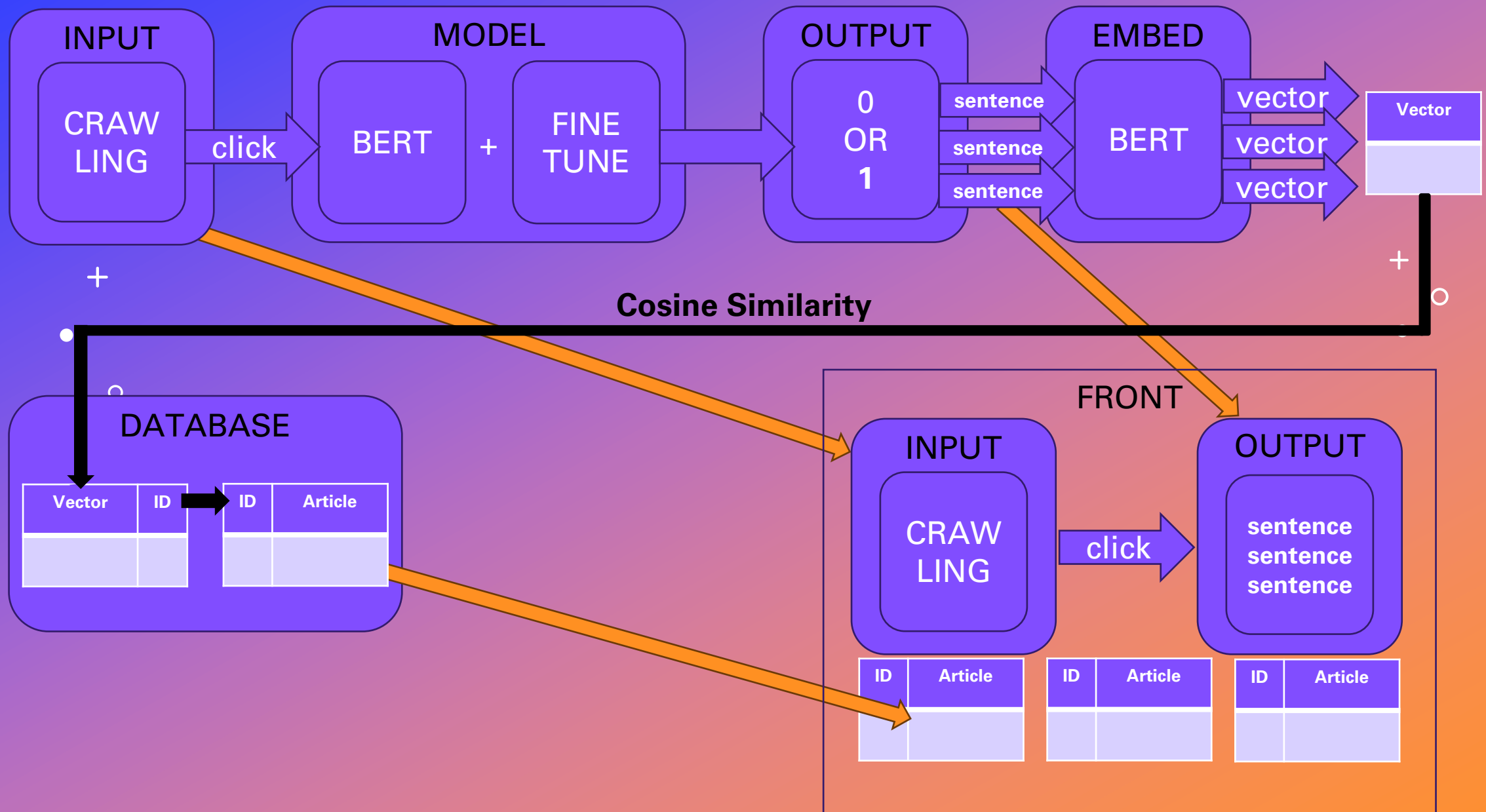
sentence	target	category
중국 투자의향기업 직접 방문·투자협약 체결 협의	0	경제
이성훈 sinawi@hanmail.net	0	경제
여수광양항만공사(사장 방희석)는 14일부터 20일까지 광양항 배후단지 투자 유치를 위해 CEO가 직접 참여하는 투자유치 활동을 중국 중남부지역에서 펼친다.	1	경제
광양만권경제자유구역청과 합동으로 진행되는 이번 투자유치활동은 방희석 사장이 직접 중국의 투자의향기업을 방문해 광양항 및 배후단지의 장점 등을 소개하고 투자협약(MOU)을 체결하는 방식으로 진행된다.	1	경제
방 사장은 먼저 광양항 서측배후단지 푸드존 투자유치를 위해 중국 운남성 소재 커피 원재료 공급업체인 운남허메이격치(주)를 방문해 광양항 배후단지 투자협약을 체결할 예정이다.	0	경제
중국의 다른 입주의향기업들도 방문해 투자유치 활동을 진행할 계획이다.	0	경제
"이번 투자유치 활동은 지난해 9월 경제청과 공사가 합동으로 중국지역 투자유치활동을 통해 투자의향기업을 선정했던 후속 조치로, 이번 방문을 통해 광양항 배후단지에 실질적인 투자가 이뤄질 수 있도록 하기 위한 것이다."	1	경제



ARCHITECTURE



전체 흐름 파악하기





ALGORITHM



어떻게 분석할 것인가?
Pre-trained Model 활용

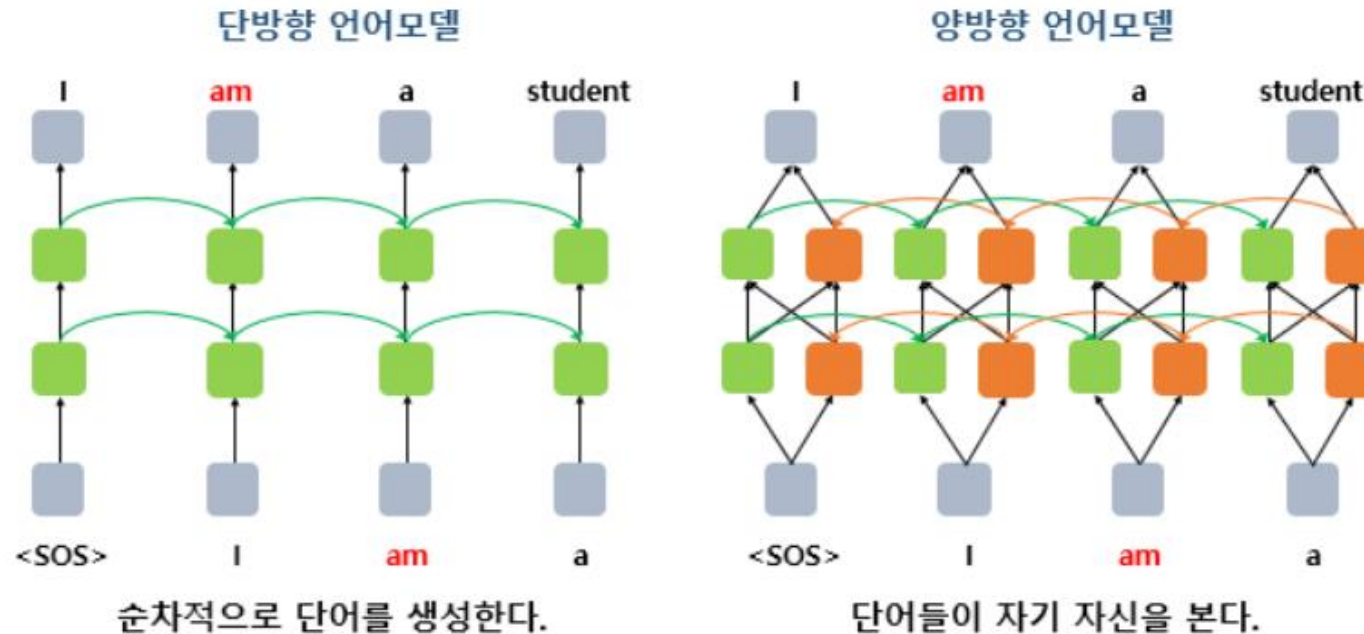
BERT

- Bidirectional Encoder Representation from Transformer(2019)
- Pre-trained 언어모델
 - 이런 학습 방식은 사람이 직접 Labeling 할 필요가 없다.
 - 대신 엄청난 양의 데이터가 필요하다.

나는 오늘 학교에 갔습니다.	
Train Data (Input)	Label
나는	오늘
나는 오늘	학교에
나는 오늘 학교에	갔습니다.

BERT

- Bidirectional Encoder Representation from Transformer(2019)
 - Bidirectional은 양방향을 의미
 - Encoder는 입력값을 숫자의 형태로 바꾸는 모듈
 - BERT는 문맥을 양방향으로 이해해서 숫자의 형태로 바꿔주는 딥러닝 모델

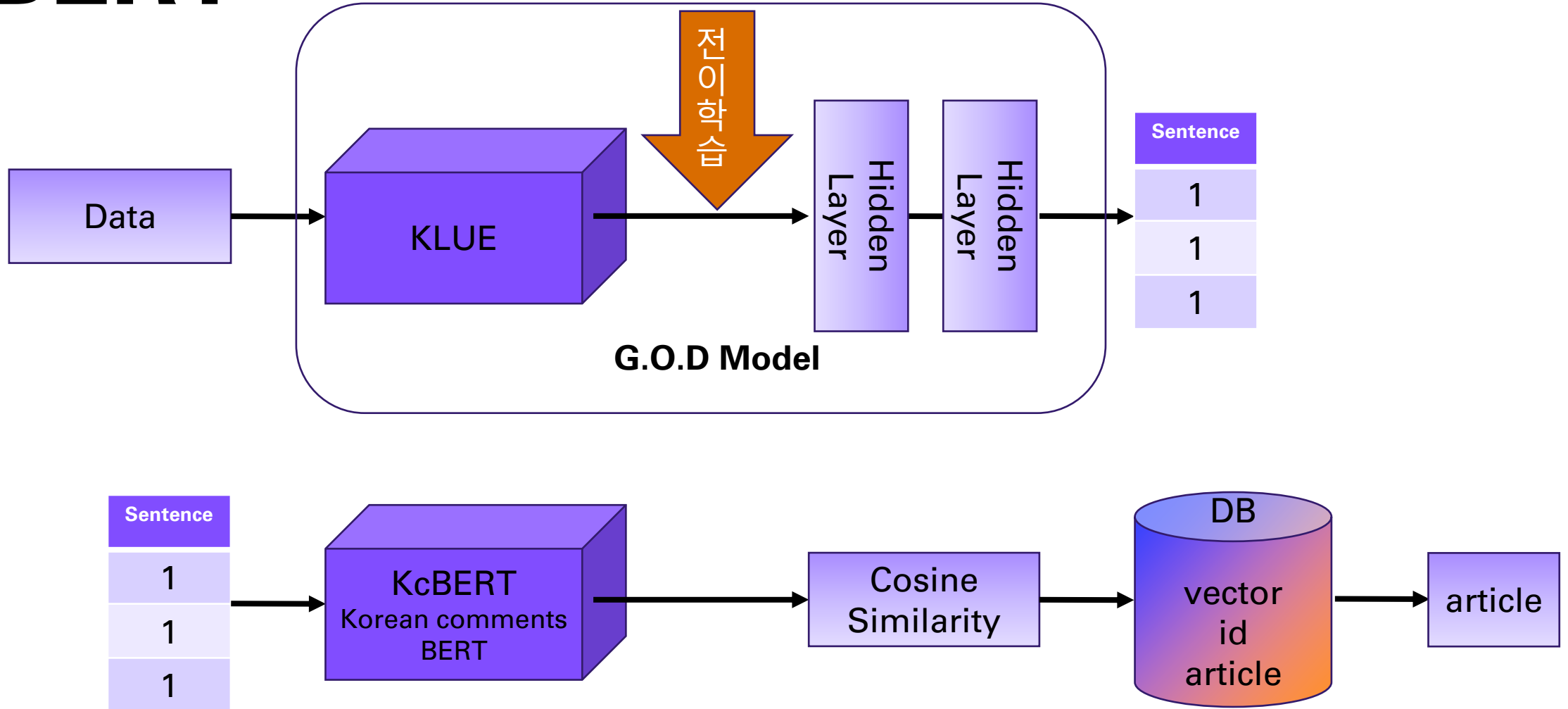


BERT를 사용한 이유

- 뉴스를 분석해서 유사한 뉴스를 추천을 해주기 위해서는 문맥을 이해해서 문장을 분석해야 된다.
- BERT는 문맥을 이해하는 Understanding 영역에서 그 성능이 뛰어나다.
- 왜냐하면 BERT는 Large-scale로 사전 학습된 언어모델이기 때문이다.
- 언어 모델이란, 일반적으로 앞 단어를 Input으로, 뒤에 나오는 단어를 Label 로 학습하는 방식을 말한다.
- 학습 방식이 단순하고 별도의 Labeling이 필요하지 않기 때문에 대용량으로 학습이 가능하다.
- 우리는 BERT를 이용하여 전이학습으로 적은 양의 Hidden Layer를 Fine Tuning하였다.
- Fine Tuning이란 적은 양의 Hidden Layer를 미세조정(재학습)하는 것을 말한다.



BERT



KLUE를 사용한 이유

- Korean Language Understanding Evaluation
- Base Multilingual Model
 - Wordpiece Tokenizer
 - annoyingly → annoying / ly
- KLUE
 - 형태소를 고려한 Tokenizing → 성능 개선





EVALUATION

SYSTEM 시연 및 성능 평가

	Layers	(learning_rate, dropout_rate)	avg_time	train_loss	train_accuracy	val_loss	val_accuracy	cnt_pred_true	diff	accuracy	
2_hidden_layers	(256, 512)	(2e-5, 0.2)	947.4s	0.0804	0.9721	0.4253	0.8769	5804	-196	0.7635	
		(2e-5, 0.3)	946.3s	0.1193	0.9577	0.3915	0.8787	5626	-374	0.7715	early_stopped
		(1e-5, 0.2)	943.5s	0.1492	0.944	0.3817	0.8692	6218	218	0.7523	early_stopped
		(1e-5, 0.3)	941s	0.1522	0.9424	0.3848	0.8701	5743	-257	0.7634	early_stopped
	(768, 256)	(2e-5, 0.2)	927.8	0.08	0.972	0.4411	0.8819	5672	-328	0.7746	
		(2e-5, 0.3)	932.3s	0.1206	0.9566	0.3923	0.8731	4712	-1288	0.7715	early_stopped
		(1e-5, 0.2)	925.8s	0.1493	0.944	0.3872	0.8658	5882	-118	0.7603	early_stopped
		(1e-5, 0.3)	98.3s	0.15	0.9444	0.3625	0.8802	6078	78	0.7755	early_stopped
	(512, 1024)	(2e-5, 0.2)	939s	0.115	0.9589	0.3756	0.879	4767	-1233	0.7784	
		(2e-5, 0.3)	940s	0.1153	0.9591	0.3835	0.8756	5408	-592	0.7707	early_stopped
		(1e-5, 0.2)	933.8s	0.0984	0.9641	0.4821	0.8595	6254	254	0.7633	
		(1e-5, 0.3)	931.3s	0.1388	0.9482	0.3744	0.8803	5098	-902	0.7755	early_stopped
single_layer	512	(2e-5, 0.2)	936.2	0.0807	0.9717	0.445	0.8797	5305	-695	0.7648	
		(2e-5, 0.3)	941.8	0.0827	0.9705	0.5088	0.8603	5944	-56	0.7576	
		(1e-5, 0.2)	937.5	0.1488	0.9451	0.3869	0.8589	6982	982	0.7413	early stopping
		(1e-5, 0.3)	934	0.0984	0.9649	0.4377	0.8637	6926	926	0.7495	
	1024	(2e-5, 0.2)	934.2	0.0827	0.971	0.4287	0.8816	5119	-881	0.7674	
		(2e-5, 0.3)	933.2	0.1458	0.9444	0.4137	0.8767	5069	-931	0.7482	
		(1e-5, 0.2)	937.25	0.1429	0.9468	0.3693	0.8729	5589	-411	0.7716	early stopping
		(1e-5, 0.3)	939.5	0.1502	0.9447	0.3914	0.8646	6966	966	0.7447	early stopping

Guardians of data

요약할 기사를 입력해 주세요.

select



doc id :

Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.



EXPECTATION

기대효과 & 활용방안

기대효과 활용방안

- 최신 기사를 요약문으로 받아볼 수 있다.
- 연관성 있는 과거의 뉴스를 함께 확인할 수 있다.
- 동일한 이슈에 대한 현재와 과거의 변화도를 파악하여 새로운 인사이트를 얻을 수 있다.

+



o



•



THANK YOU

Q & A