

[Web발신]

성북구 플코스로 모십니다

NLP 챗봇 2팀
민윤기, 최규빈, 홍여빈



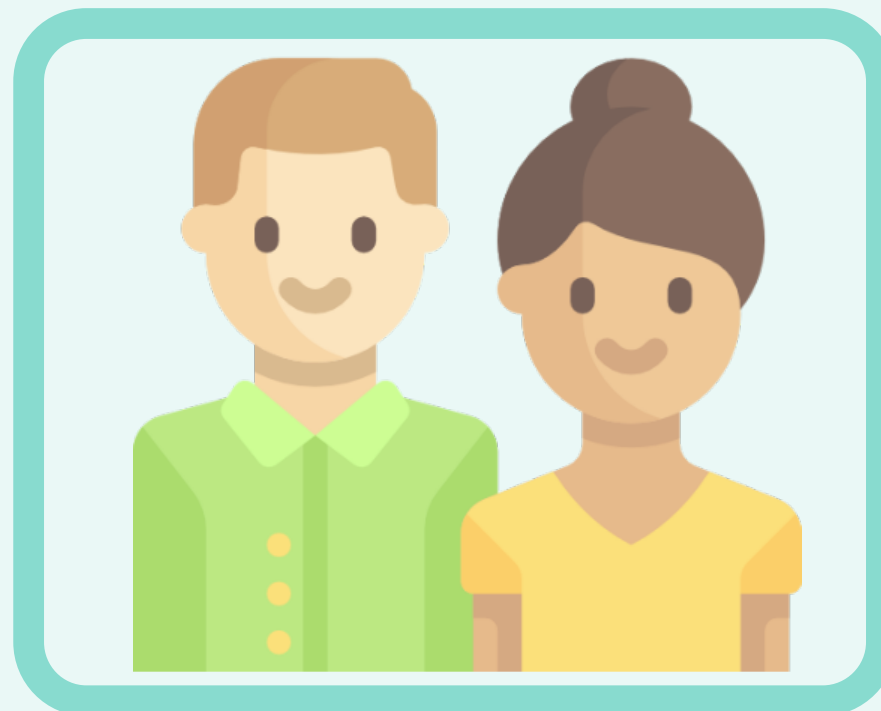
성북구에서 데이트하는 젊은이들을 위한
데이트코스 추천 챗봇

맛집



36곳

데이트코스



51곳

술집



26곳

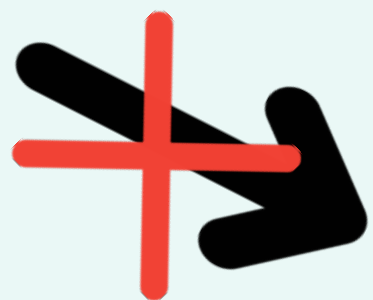
STS (Semantic Textual Similarity)

1. 문장을 어떤 기준으로 분리할 것인가(토큰나이징)
2. 각 토큰을 어떻게 숫자로 바꿀 것인가(워드 임베딩)
3. 두 문장은 어떻게 비교할 것인가(유사도 측정)

달콤한 음식이 먹고싶어



오늘은 달달한게 땡기네



오늘은 얼큰하고 싶어

챗봇이란?

트랜잭션 챗봇



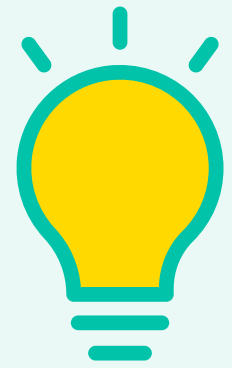
작업 지향적, rule-based
ex) 은행 챗봇, 어흥톡

대화형 챗봇



상호작용, 정교한 인터랙티브
ex) ChatGPT

트랜잭션 챗봇 PORORO



01 pororo

kakao brain에서 개발한 한국
어 종합 자연어처리 라이브러리

03 토큰나이징

Mecab 형태소 분석기
Okt, Kkma, Komoran 등의
다른 분석기보다 빠른 속도

02 사용 분야

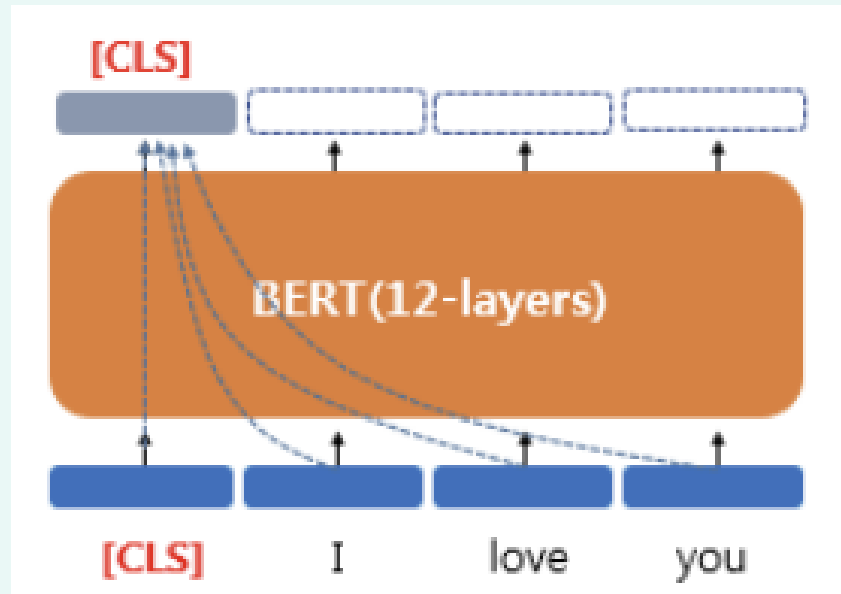
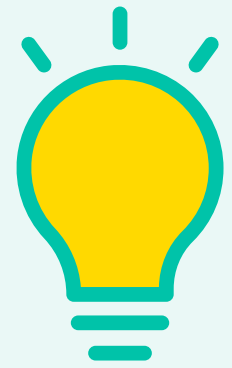
토큰나이징, 워드임베딩부터 문서
요약, 맞춤법 교정, 번역, 채점, 빈
칸 채우기, 이미지 캡션 등

04 센텐스 임베딩

SBERT 기반
(Sentence - BERT)

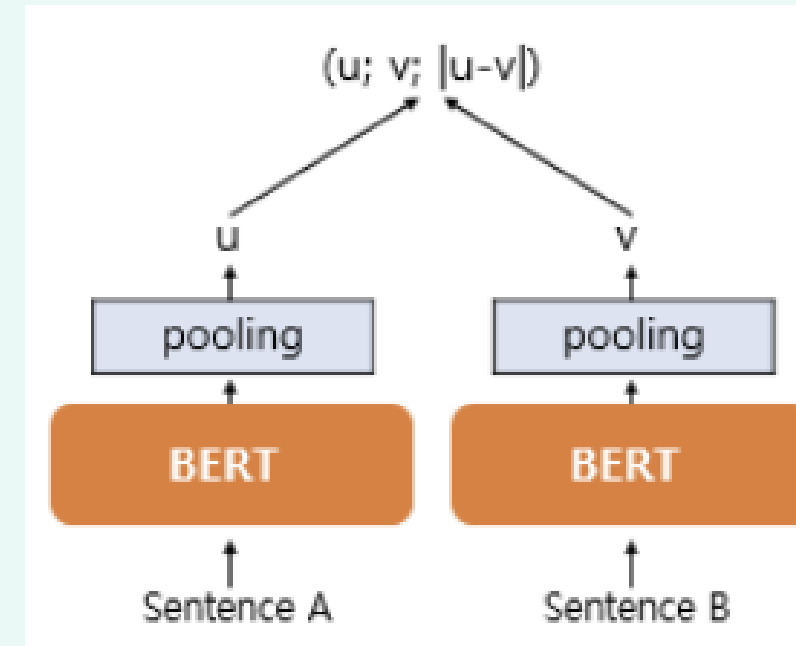
트랜잭션 챗봇

SBERT (Sentence-BERT)



BERT

"나는 밥을"이라는 문장을
받았을 때 (먹는다) 라는
단어를 생성해낼 수 있는가



SBERT


"나는 밥을 먹는다"라는
문장과 "나는 식사를 한다"
라는 문장이 유사하다고 인
지할 수 있는가

대화형 챗봇 KoGPT



 `skt/kogpt2-base-v2`

 • Updated Sep 24, 2021 • ↓ 310k • ♥ 15

 `skt/ko-gpt-trinity-1.2B-v0.5`

 • Updated Sep 24, 2021 • ↓ 1.61k • ♥ 20

01 kogpt2

gpt2 기반 디코더 언어모델
다른 한국어 gpt 모델에 비해 비
교적 가벼운 모델

02 ko-gpt-trinity

gpt3의 구조를 모방한 디코더 언어
모델, kogpt2보다 10배가량 많은
파라미터로 복잡한 텍스트를 더 잘
이해하는 모델

03 토큰나이징&워드임베딩

SPM 모델을 통해 토큰나이징과 워드임베딩을 별
도로 진행하지 않고 하나의 과정을 묶음

대화형 챗봇

SPM(Sentence Piece Model)



OOV(Out Of Vocabulary) 해결의 열쇠

단어의 형성			
단어	단일어		
	복합어	합성어	
		파생어	접두사+어근
			어근+접미사

귀엽다, 귀여미, 귀요미, 귀염둥이



'귀' + 'ㅡ'



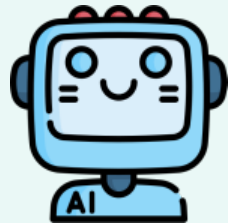
귀염둥이

BPE(Byte Piece Encoding)

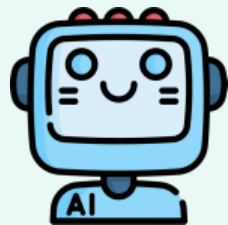
+

Unigram Language Model

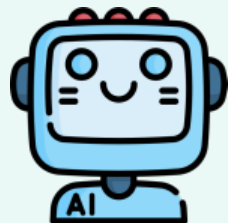
실제 대화 [지역 선택]



반가워요 저는 당신의 성북구 데이트 코스를 짜줄 성북FC예요.



우선 안암 / 혜화 / 성신 / 성북구 전체 중 어디서 데이트하고 싶은지 말해주세요.



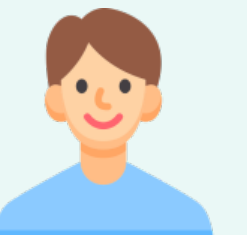
'안'에서 데이트하고 싶으신 것 맞나요?

핵심 키워드

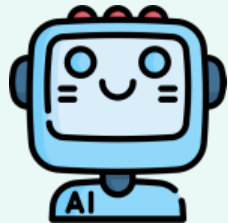
내일 안에서 놀고싶어!



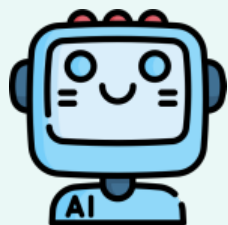
응 맞아!



실제 대화 (식당 선택)

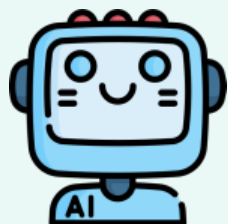


어떤 식당에 가고 싶은지 말씀해주세요!



썬더치킨 어떠신가요?

'닭 오리오리 후라이드 양념치킨 치킨 전문점'을 팔고
'요즘같은 고물가...사장님 친절'의 리뷰도 받았어요

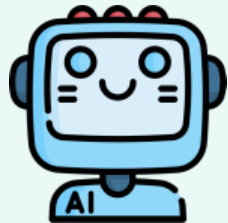


여기를 추천하는 이유는 치킨 맛이 굉장히 맛있다는 점입니
다. 다만 <썬더치킨>에서 치킨을 먹었을 때 아쉬운 점이 있
다면 생맥주 가격이 6000원이다.

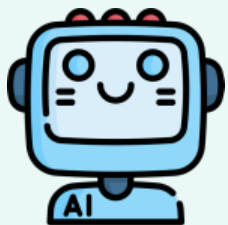
내일은 치킨을 좀 먹고싶네



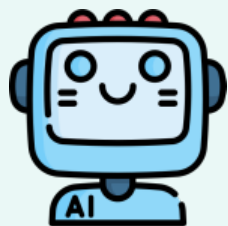
실제 대화 (데이트코스)



점심먹은 후에는 어떤 데이트장소에 가고
싶으신가요?



제 생각에는 낙산공원에 가는 게 좋은 것 같아요
낙산공원은 활동적으로 산책하기 좋은 공원입니다

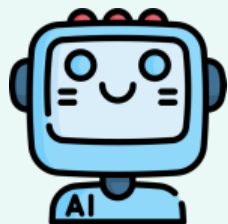


산책이란 무엇인가를 고민하다 보면 산책의 의미 자체가
'자연스럽게, 산책하면서'까지 가도록 설계되어 있기 때문
입니다. 그러다 보니 걷다 보면 다른 세상이 보입니다. 그 세
상에는 또 다른 세상이 있습니다.

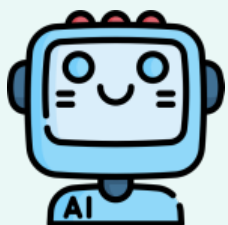
점심 먹고 가볍게 야외에서
산책하고 싶어



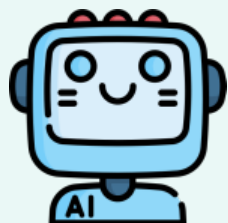
실제 대화 (술집 선택)



오늘 방문하고자 하는 술집의 원하는 분위기, 가격대 등이 있으면 말해주세요.



56st는 어떠신가요?
'식사료, 가성비, 조용함, 데이트'의 특징이 있어요

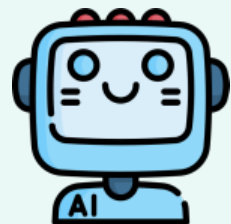


여기를 추천하는 이유는 다른 곳과는 다르게, 가격도 저렴하고 분위기도 좋아서 식사시간대에도 사람이 많아서 식사를 하고나면 여유롭게 식사를 할 수 있어서 좋습니다.

대화하기 좋은 조용한 분위기의
가게에 가고싶어

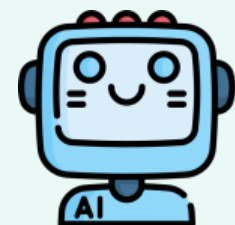
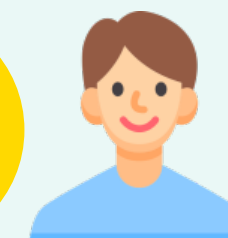


실제 대화 (마무리)



마음에 드셨나요?

네 완벽해요



제가 완전하죠

시사점 & 한계점

KEYWORD

A

문장의 부자연스러움

리뷰 데이터를 가져와서 부드럽게
재편집하지 않고 그대로 붙이다보
니 부자연스러운 어투가 됨

KEYWORD

B

RAM 문제

현재 가장 강력한 챗봇 모델인
kakao brain의 KoGPT는
Colab의 기본 제공 RAM을 초과
해 RAM을 구매해야함

KEYWORD

C

감성 분류

유사도 기반 텍스트 분석은 좋은 성
능을 보였으나 일상대화의 맥락 읽
기는 여전히 부족하므로 감성분류
가 결합된다면 더 좋을듯함

THANK
YOU