

최종 프로젝트 기획안

2023년 5월 4일

프로젝트 조명	매추리 (매일을 간추리다)
프로젝트 팀원	팀명: 매추리 팀장: 조민우 팀원: 고은영, 박기현, 유재찬, 윤승욱, 조민우, 하정수
프로젝트 주제	바쁜 출근길 직장인들을 위한 오간뉴 (오늘의 간추린 뉴스)
프로젝트 수행 방향 (주요 기능 설명)	<p>본 프로젝트의 주요기능은 바쁜 출근시간대의 직장인들에게 간추린 뉴스 제공입니다.</p> <p>특정 시간대의 뉴스 데이터를 수집하고 저장한 뒤, 딥러닝 기반 자연어 처리 모델을 활용하여</p> <p>뉴스를 요약하고 웹서비스 통해 제공합니다. 뉴스는 사용자가 선택한 항목(주제, 언론사 등)을 기반으로 편성하여 기사를 요약한 뒤 제공됩니다.</p> <p>기능 구현을 위해 기반이 된 모델은 사전학습을 마친 언어 모델 위에 작은 모듈을 조금 더 쌓아 태스크를 수행하는 구조의 트랜스 포머 계열 언어 모델, BART입니다.</p> <p>BART 모델은 BERT모델과 다르게 인코더 및 디코더 레이어 구조를 가진 seq2seq모델로 텍스트 형태의 결과물을 반환합니다.</p> <p>본 프로젝트에서는 한국어 성능 한계를 극복하기 위해 KOBART모델을 기반으로 학습한 KOBART -base-v2모델을 활용합니다.</p> <p>전이 학습 및 파인튜닝 과정을 거쳐 여러 모델을 비교 후, 최종적으로 base v2모델을 서빙모델로 선정하였습니다.</p>
프로젝트 수행 도구	<ul style="list-style-type: none">- Github : 소스코드 형상 관리- Visual Studio Code : 코드 작성 및 실시간 공유- Colab : 코드 작성 및 모델 학습- AWS : 모델 학습 및 프로젝트 배포- Notion : 프로젝트 이슈 관리

- 구글 드라이브 : 문서 공유

프로젝트 목적	AI가 미디어의 생산 및 유통과정에 주도적으로 참여하게되어, AI솔루션(Transformer)을 활용하여 변형한 콘텐츠를, 이용자가 선택한 성향에 맞춘 개인화 서비스와 함께 제공
필수 기능	<ul style="list-style-type: none">- 자연어 처리 모델을 활용한 뉴스 요약 제공- 뉴스 기사 데이터 크롤링 및 전처리- 웹 기반 인터페이스를 통해 서비스 제공- 회원이 선택한 항목에 맞춘 뉴스 편성
포함 기술	웹 크롤링, NLP, Pytorch, Django, Javascript, HTML, CSS, MariaDB, AWS, Python, matplotlib, seaborn

□WBS 첨부(OR 별도로 폴더에 업로드)