

캡스톤디자인 중간보고서

제 목	국문	LLM 기반 한국어 문법 오류 교정 서비스		
	영문	LLM-based Korean Grammatical Error Correction Service		
진 행 상 황	준요마일스톤	<ul style="list-style-type: none">- 인공지능 모델을 기반으로 한 한국어 GEC, AES 실태 조사- Transformer Decoder LLM 기반 지식 학습- 한국어에 특화된 llama2 기반 BLLOSSOM 모델 STF 학습 코드 작성- GEC Annotation Tool인 KAGAS의 결과물 분석- GEC 평가에 사용되는 m2 scorer, gleu 분석 및 사용 코드 작성- AES 평가에 사용되는 MSE, AWK 분석 및 사용 코드 작성- 기존 모델과의 GEC 성능 비교 및 모델의 성능 평가- L1, L2, L1+L2 세 데이터 셋에 대해 모델 학습 및 성능 비교- AES에 대한 모델 학습 및 성능 확인- 모델을 반영한 시스템 인터페이스 설계- 모델의 추론 파이프라인 설계 및 코드 작성- 모델 서빙용 추론 서버 개발		
	진행상황	<ul style="list-style-type: none">- Transformer Decoder LLM 기반 지식 학습- 한국어에 특화된 llama2 기반 BLLOSSOM 모델 STF 학습 코드 작성- GEC Annotation Tool인 KAGAS의 결과물 분석- GEC 평가에 사용되는 m2 scorer, gleu 분석 및 사용 코드 작성- AES 평가에 사용되는 MSE, AWK 분석 및 사용 코드 작성- 기존 모델과의 GEC 성능 비교 및 모델의 성능 평가- L1, L2, L1+L2 세 데이터 셋에 대해 모델 학습 및 성능 비교- AES에 대한 모델 학습 및 성능 확인- 초기 단계 모델 서버 개발		
산출물	요구사항 정의서(별첨 1), 중간보고서(별첨 2)			
팀 구성원	학년	학 번	이 름	연락처(전화번호/이메일)
	4	20191120	이용빈	010-6560-9717 / 20191120@edu.hanbat.ac.kr
	4	20191780	육정훈	010-4713-9845 / 20191780@edu.hanbat.ac.kr
컴퓨터공학과의 프로젝트 관리규정에 따라 다음과 같이 요구사항 정의서와 중간보고서를 제출합니다				
2024 년 5 월 23 일				
책임자 : 이용빈 (인)				
지도교수 : (인)				

[별첨1]

프로젝트명 : LLM 기반 한국어 문법 오류 교정 서비스

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):이용빈
육정훈

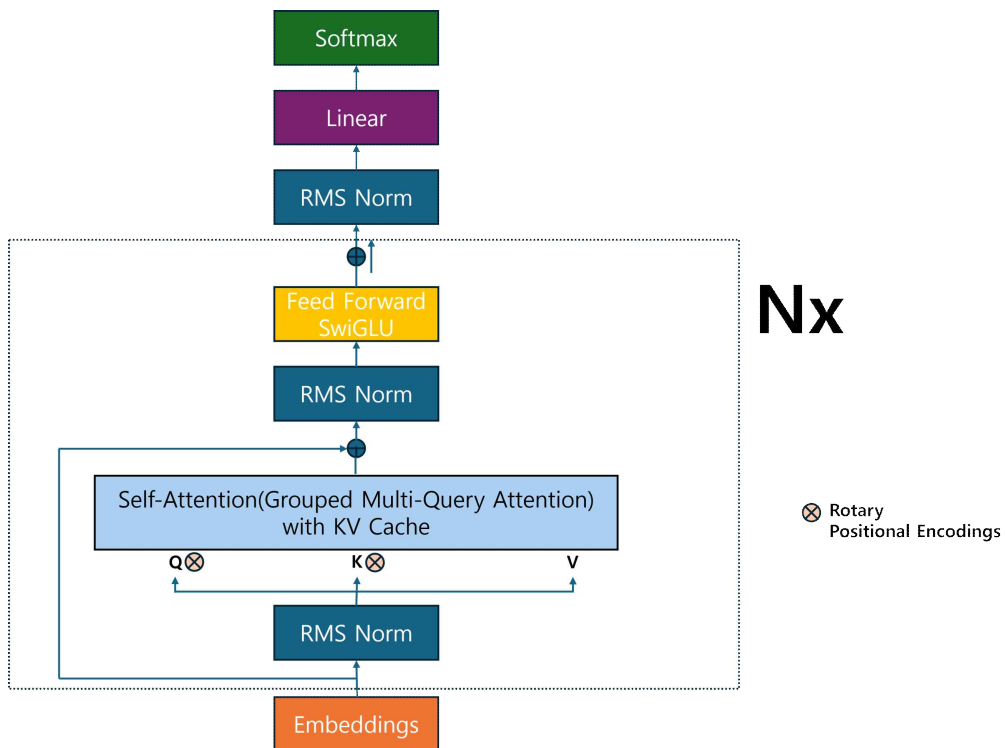
대표 연락처:010-6560-9717
e-mail: 20191120@edu.hanbat.ac.kr

목차

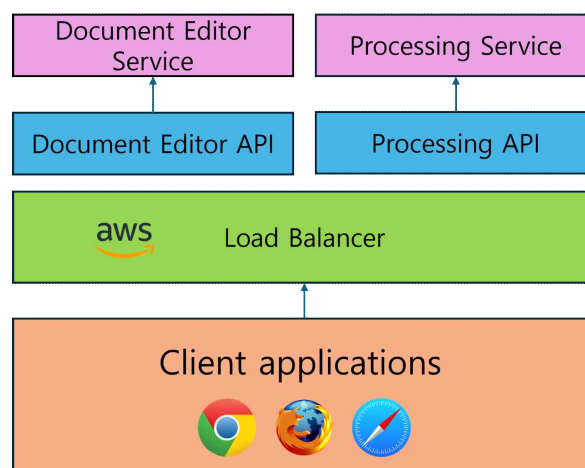
1. 개요
2. 시스템 장비 구성요구사항
3. 기능 요구사항
4. 성능 요구사항
5. 인터페이스 요구사항
6. 데이터 요구사항
7. 테스트 요구사항
8. 보안 요구사항
9. 품질 요구사항
10. 제약 사항
11. 프로젝트 관리 요구사항

1. 시스템 개요

- LLaMa 기반의 글쓰기 평가 및 문법 오류 교정 거대 언어 모델을 활용한 웹 서비스



LLaMa 모델 아키텍처



예상 서비스 아키텍처

2. 시스템 장비 구성요구사항

요구사항 고유번호		ECR-001
요구사항 명칭		장비 요구사항
요구사항 분류		시스템 장비 구성 요구사항
요구사항 상세 설명	정의	모델 학습 장비
	세부 내용	<ul style="list-style-type: none"> - 장비 품목: GPU(NVIDIA A6000) - 장비 수량: 2개 - 장비 기능: 모델의 연산 속도를 높인다. - 장비 성능 및 특징: 메모리 40GB - 시간 제약사항: 다른 팀과 동일한 GPU 사용으로 할당 시간을 정해서 사용 - 자원 제약사항: 해당 사항 없음 - 장애 처리: 해당 사항 없음

3. 기능 요구사항

요구사항 고유번호		SFR-001
요구사항 명칭		AI 모델 기반 서비스 개발
요구사항 분류		기능
요구사항 상세 설명	정의	GEC 어플리케이션 시스템 개발
	세부 내용	<ul style="list-style-type: none"> - 외국인이 작성한 한국어 문법 오류 유형 데이터에 대해 한국어에 특화된 LLM 모델을 학습시켜 표현 정정을 학습 시킬 것 - 한국이 작성한 한국어 문법 오류 유형 데이터에 대해 한국어에 특화된 LLM 모델을 학습시켜 표현 정정을 학습 시킬 것 - 위의 두 가지 오류 유형 혼합 데이터에 대해 추가적으로 학습하여 두 유형 모두에서 잘 작동하는 모델을 학습할 것 - 모델이 전반적인 글쓰기 수준에 대한 평가를 진행할 수 있도록 글쓰기의 수준 및 점수를 학습 시킬 것 - 클라우드 서버 시스템으로 사용자의 형태와 규모에 상관 없이 원활한 서비스 운영이 가능하도록 할 것

4. 성능 요구사항

요구사항 고유번호		PER-001
요구사항 명칭		처리 속도 및 시간
요구사항 분류		성능 요구사항
요구사항 상세 설명	정의	처리 속도 및 시간
	세부 내용	- 모델이 입력된 오류 문장에 대해 정정된 문장을 출력하는 시간을 의미함 - 모델이 글쓰기에 대해 평가하고 점수를 출력하는 시간을 의미함

5. 인터페이스 요구사항

요구사항 고유번호		SIR-001
요구사항 명칭		문서 편집 인터페이스
요구사항 분류		사용자 인터페이스
요구사항 상세 설명	정의	서비스 대상 문서 편집기
	세부 내용	- 글쓰기 창에는 사용자가 자유롭게 글쓰기를 수행할 수 있도록 인터페이스 구현 - 사용자의 입력 창과 글쓰기 점수, 오류 정정 제안 출력 창을 구분하여 인터페이스에 표시

6. 데이터 요구사항

요구사항 고유번호		DAR-001
요구사항 명칭		데이터 표준 준수
요구사항 분류		데이터
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none">- 연구 기관에서 인계받은 L1, L2, L1+L2 GEC 학습 데이터 사용- L1에 대한 오류 문장과 글쓰기 평가 데이터 셋 사용

요구사항 고유번호		DAR-002
요구사항 명칭		초기자료 구축
요구사항 분류		데이터
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none">- 학습에 올바른 사용을 위해 초기 데이터 구축 및 GEC, AES 데이터를 tsv로 저장

7. 테스트 요구사항

요구사항 고유번호		TER-001
요구사항 명칭		성능 테스트
요구사항 분류		테스트
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none">- 학습이 끝난 모델에 대해 원하는 결과(오류가 해결된 문장)가 제대로 generation 되는지 테스트하고 점검하기 위한 평가 기준으로 m2 score, gleu 사용- 기존 존재하는 한국어 GEC 모델과의 성능 비교

요구사항 고유번호		TER-002
요구사항 명칭		성능 테스트
요구사항 분류		테스트
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none">- 글쓰기에 대해 올바른 수준 평가와 점수를 출력하는지 AWK, ACC, MSE로 평가

8. 보안 요구사항

요구사항 고유번호		SER-001
요구사항 명칭		보안지침 준수
요구사항 분류		보안
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none"> - 데이터 셋의 라이선스 관련 이용 정책과 공개 여부에 유의하여 개발이 수행되어야 함 - 모델의 공개 여부에 유의하여 개발이 수행되어야 함

9. 품질 요구사항

요구사항 고유번호		QUR-001
요구사항 명칭		데이터 품질 관리
요구사항 분류		품질
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none"> - 학습 데이터에는 중복이 존재하지 않아야 하며, 정제 과정을 통해 제거해야 한다. - GEC 데이터 및 AES 데이터는 모델에 대해 정형화된 출력이 생성될 수 있도록 다양한 값의 데이터가 존재해야 한다. - 학습 데이터 및 검증, 테스트 데이터는 학습에 방해되지 않도록 결측치 및 중복 여부, 데이터 자료형 등을 파악하여 제외 혹은 포함 여부를 판단하여 사용하여야 한다. - 글쓰기 수준의 카테고리는 객관적인 기준에 따라 납득 가능한 범위여야 하며, 모호성이 존재하지 않는다. - 각 데이터는 클래스에 대해 편향되지 않게 적절한 비율을 유지해야 한다.

10. 제약 사항

요구사항 고유번호		COR-001
요구사항 명칭		시스템 개발과 설계 및 구현 제약사항
요구사항 분류		제약사항
요구사항 상세 설명	세부 내용	<ul style="list-style-type: none"> - 현재 보유하여 활용 가능한 H/W, S/W를 최대한 활용함. - 대부분의 인공지능 모델 개발에 사용되는 Python, PyTorch를 사용함. - 대규모 LLM의 사용을 위해, huggingface에서 제공하는 transformers 패키지를 사용함. - 데이터의 오류 형태 및 위치를 확인할 수 있는 KAGAS를 통해 annotation을 생성하며, 이를 이용하여 m2 score를 이용해 전반적인 성능 도출

11. 프로젝트 관리 요구사항

요구사항 고유번호		PMR-001
요구사항 명칭		품질 관리
요구사항 분류		프로젝트 관리
요구사항 상세 설명	세부 내용	<p>세부 작업 분할 구조</p> <ol style="list-style-type: none"> 분석 <ul style="list-style-type: none"> - KAGAS, m2 scorer, gleu score 사용 기법 조사 - 한국어에 특화된 대규모 LLM 탐색 시스템 설계 <ul style="list-style-type: none"> - GEC, AES 데이터 셋과 한국어에 특화된 LLM을 이용한 SFT 구현 실험 <ul style="list-style-type: none"> - L1, L2, L1+L2 데이터 셋을 이용하고 한국어에 특화된 LLM을 활용한 모델을 사용하여 GEC, AES 모델 학습 - MSE, AWK, m2 score, gleu score, acc를 이용한 성능 평가 <p>- 프로젝트 수행조직에 대한 구성, 역할</p> <p>이용빈: 논문 및 자료조사, mlops 구현 및 시스템 개발</p> <p>육정훈: 논문 및 자료조사, 모델 코드 작성 및 실험</p>

[별첨2]

중간보고서

1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 분석, 설계, 구현(소스코드 작성) 및 테스트한 내용을 기술하시오.

- 인공지능 모델을 기반으로 한 한국어 GEC, AES 실태 조사
- Transformer Decoder LLM 기반 지식 학습
- 한국어에 특화된 llama2 기반 BLLOSSOM 모델 STF 학습 코드 작성
- GEC Annotation Tool인 KAGAS의 결과물 분석
- GEC 평가에 사용되는 m2 scorer, gleu 분석 및 사용 코드 작성
- AES 평가에 사용되는 MSE, AWK 분석 및 사용 코드 작성
- 기존 모델과의 GEC 성능 비교 및 모델의 성능 평가
- L1, L2, L1+L2 세 데이터 셋에 대해 모델 학습 및 성능 비교
- AES에 대한 모델 학습 및 성능 확인

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.

- 한국어 GEC 및 AES 인공지능 모델의 현황 조사

결과: 시장 및 학계의 기존 접근 방법과 기술을 조사하여 프로젝트의 기초를 마련하였음.
Transformer 기반 모델과 비교하여 LLM의 잠재력을 파악하였음.

- Transformer Decoder 및 LLM 기반 모델의 학습

결과: Transformer Decoder와 LLM을 활용하여 한국어 GEC 및 AES 모델을 개발하였음.
BLLOSSOM 모델을 사용하여 한국어에 최적화된 학습을 시도하였음.

문제점: 기존의 트레이닝 데이터셋과 호환성 문제가 발생하였음.

해결방안: 데이터 포맷을 조정하고 전처리 단계를 추가로 설정하여 데이터 셋과 모델 간의 호환성 해결함.

- 팀원의 책임 및 역할 수행에 대한 결과

결과: 모든 팀원이 기술 조사, 모델 개발, 결과 분석 등의 역할을 효과적으로 수행하였음.

캡스톤 디자인 I 중간보고서 채점표

평가도구	평 가 항 목	평 가 점 수				
		1	2	3	4	5
중간 보고서 및 실행 결과	1. 요구사항 정의서(기능, 성능, 인터페이스 등)가 구체적으로 작성되었는가?					
	2. 요구분석, 설계 산출물(모델, 프로토타입 등)의 내용이 충실한가?					
	3. 설계 및 구현 문제를 위해 적용한 이론, 문제해결 방법이 제시되었으며 그 적용이 적합한가?					
	4. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)가 버그 없이 실행되었는가?					
	5. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)의 성능 요구사항은 충족되었는가?					
도구활용	6. 설계 및 구현을 위해 도구가 적절히 활용되었는가?					
	7. 도구의 활용수준(능숙도)은 프로젝트 수행에 적합한가?					
팀원의 업무 및 역할	8. 팀원의 업무분담에 따른 역할 및 협력이 충실히 이루어졌는가? (평가자에 의한 질의)					
	9. 프로젝트 중간 진척상황에 대해 팀원이 충분히 인지하고 있는가?(평가자에 의한 질의)					
합계						
*검토 의견(최종완료 때까지 보완해야할 점에 대해 작성 요망)						
심사위원(소속):		(이름)			(인)	