

2020 졸업프로젝트 :

Knowledge-Based Dataset에 기반한 질의응답 챗봇



001 개요

- 프로토타입 이미지



002 요구사항 분석

- Knowledge-Based Database 구축 기능 완성
- 질의응답 챗봇 구현
- SQuAD 정확도 70% 달성
- 처리시간 단축



003 아키텍처 다이어그램

- 인터페이스 정의 및 요구사항 맵핑



003 시스템 테스트 케이스



201410349 서원준



201411310 장승원



201610099 이수민



001 개요

프로토타입 이미지

프로젝트 구현 목표

- KBLRN : End-to-End Learning of Knowledge-Base Representations with Latent, Relational, and Numerical Features (2017년 논문)
- 데이터베이스는 (Object, Relation, Entity)로 구성됨 (I, Love, You) (Adam, Drinks, Water)
- 사용자가 입력한 문장을 분석해, 정보(Relations)를 추출하고 데이터베이스에 저장
- 사용자가 질문을 입력하면 데이터베이스를 참조해 알맞은 대답을 반환



<사용자가 상호작용하게 될 응용 프로그램의 프로토타입 이미지>



002 요구사항 분석

기능
요구사항

1. Knowledge Based Database 구축 기능 완성

1.1 입력 문장에서 Entity - Relation - Entity의 관계를 추출

- 입력 받은 데이터를 전처리하여 모델을 돌린 후 KB 데이터 형식에 맞게 변형

1.2 KB 데이터베이스 구축

- (Entity, Relation, Entity) 의 튜플을 CSV 파일에 누적시켜 저장

2. 질의응답 챗봇 구현

2.1 사전 구성된 KBDB 를 참조해 질문에 대한 대답을 반환

- (I, Love, ?) -> (I, Love, You). Who, When, Where, What, Why, How 형태의 질문 권장

2.2 사용자와 상호작용할 수 있는 프로그램 작성

- 로컬 환경에서 구동, PyQt로 UI 작성.

3. SQuAD 정확도 70% 달성

- 질의응답 데이터셋 Stanford Question Answering Dataset을 이용해 성능 측정

4. 처리시간 단축

- 학습 시: 문장당 처리시간 3초 이내 목표
- 질문 시: 응답시간 1초 이내 목표

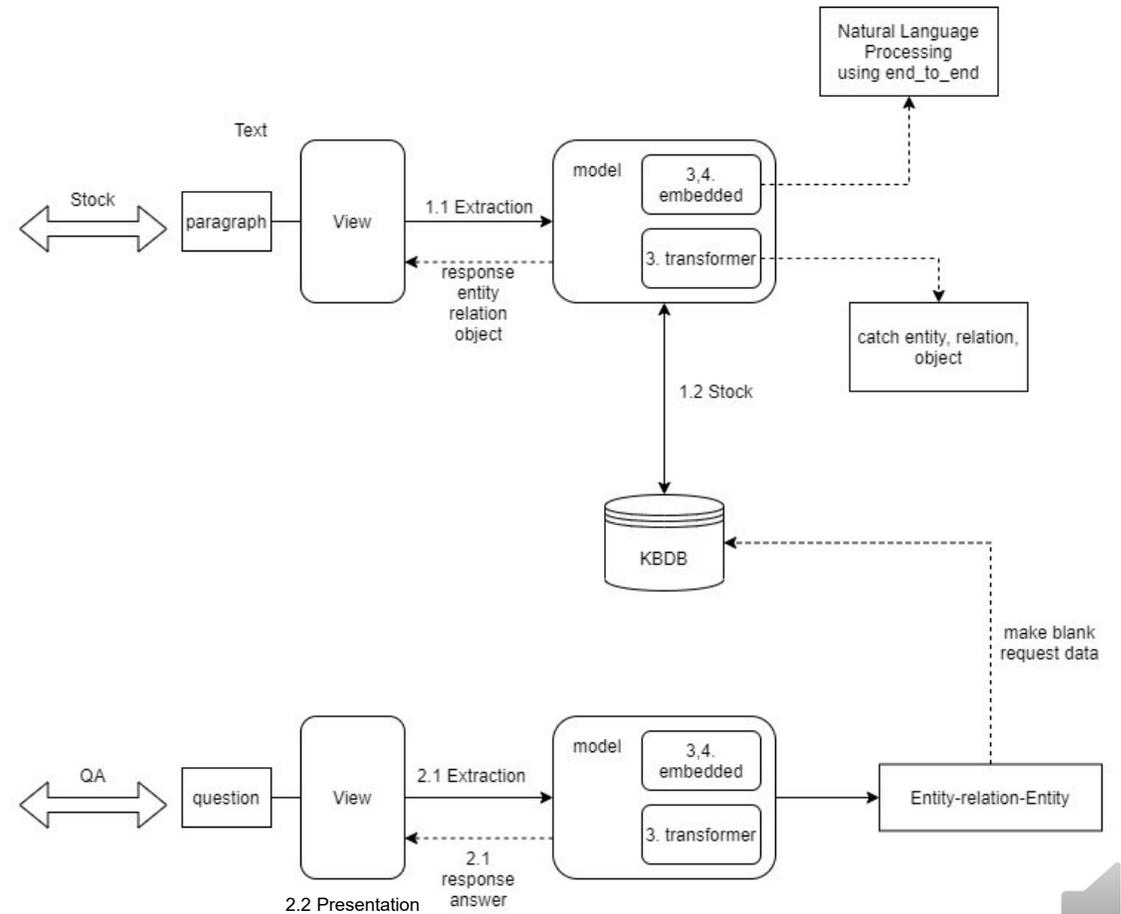
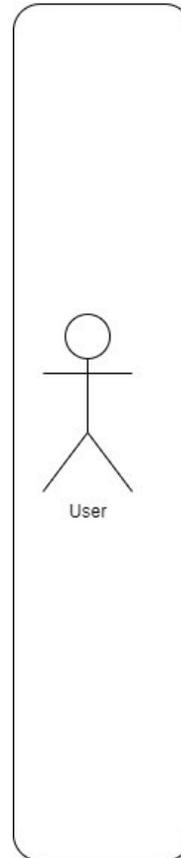
비기능
요구사항



003 아키텍처 다이어그램

인터페이스 정의 및 요구사항 맵핑

- 1.1. Extraction : 입력 데이터에서 (Entity, Relation, Entity) 추출
- 1.2. Stock : 추출한 튜플을 누적시켜 데이터베이스 구축
- 2.1. Response-Answer : KBDB를 참조해 질문에 응답
- 2.2 Presentation : 사용자에게 정보 표시
- 3. Transformer : 정확도 개선을 위해 투입
- 4. Embedded : 처리시간 단축을 위해 여러 가지 Embedding 기법 사용 시도



004 시스템 테스트 케이스

Case No.	테스트케이스 목표	입력 상황	예상 결과
TEST 1.1 (기능)	입력 문장으로부터 Entity-Relation-Entity 관계가 성공적으로 도출되어야 함	아메리카노의 가격은 4000 원이다.	(아메리카노, 가격, 4000 원) 라는 관계가 도출됨
TEST 1.2 (기능)	입력 문장의 (Entity-Relation-Entity) 튜플이 csv 파일에 누적되어 저장되고, KBDB 를 구축	아메리카노의 가격은 4000 원이다.	기존 csv 파일에 (아메리카노, 가격, 4000 원) 튜플이 추가됨
TEST 2.1 (기능)	구축된 KBDB 를 참조해 입력한 질문에 대한 대답을 반환함	아메리카노가 얼마인가요?	아메리카노는 4000 원 입니다.
TEST 2.2 (기능)	사용자 UI 를 통해 KBDB 와의 상호작용이 이루어져야 함	프로그램에 학습 데이터를 투입 혹은 질문	데이터베이스에 데이터 누적, 질문에 대한 답변 반환

Case No.	테스트케이스 목표	입력 상황	예상 결과
TEST 3 (비기능)	SQuAD 테스트 데이터셋 기준으로 정답률 70% 이상 달성	SQuAD 의 질의응답 테스트 데이터를 개발한 모델에 입력	테스트 결과 정답률 85%
TEST 4 (비기능)	학습 시 문장당 3 초, 질의응답 시 1 초의 처리시간을 달성	문장 입력 : 아메리카노의 가격은 4 천원이다. 질문 : 아메리카노는 얼마인가?	학습 소요 시간 2.3 초, 응답 소요 시간 0.8 초

