

2020 졸업프로젝트 :

## 한국어 Simple Question의 SPARQL 변환

- 001 요구사항 분석
- 002 시스템 테스트 케이스
- 003 상위 디자인
- 004 상세 디자인
- 005 Traceability Matrix



201410349 서원준



201411310 장승원



201610099 이



# 001 요구사항 분석

가능  
요구사항

## 1. 한국어 Simple Question to SPARQL을 통한 질의응답

### 1.1 Sparql 데이터로 변환하기 위해 AI 구축

1.1.1 Sparql 데이터의 Subject를 추출하기 위하여 Transformer 구축

1.1.2 Sparql 데이터의 Predicated를 추출하기 위하여 Convolutional Neural Network (CNN) 구축

### 1.2 질문 데이터 형태소 분석

1.2.1 Simple question을 한국어로 입력

1.2.2 입력받은 문장을 형태소 형식으로 분해

### 1.3 질문 데이터를 Sparql로 변형 및 정답 찾기

1.1.3. 형태소로 분해된 데이터에서 1.1.1과 1.1.2에서 구축된 AI로 Subject, Predicated 추출

1.1.3. Subject, Predicated를 SPARQL로 합성

1.1.3. SPARQL을 DBPedia에 참조하여 답변 찾기

비가능  
요구사항

## 2. Simple Question 정확도 70% 달성

- Nl to SQL 정확도 70% 달성
- DBPedia 데이터를 이용한 질의응답 정답률 70% 달성
- 달성하지 못할 경우 최적의 방법을 찾기



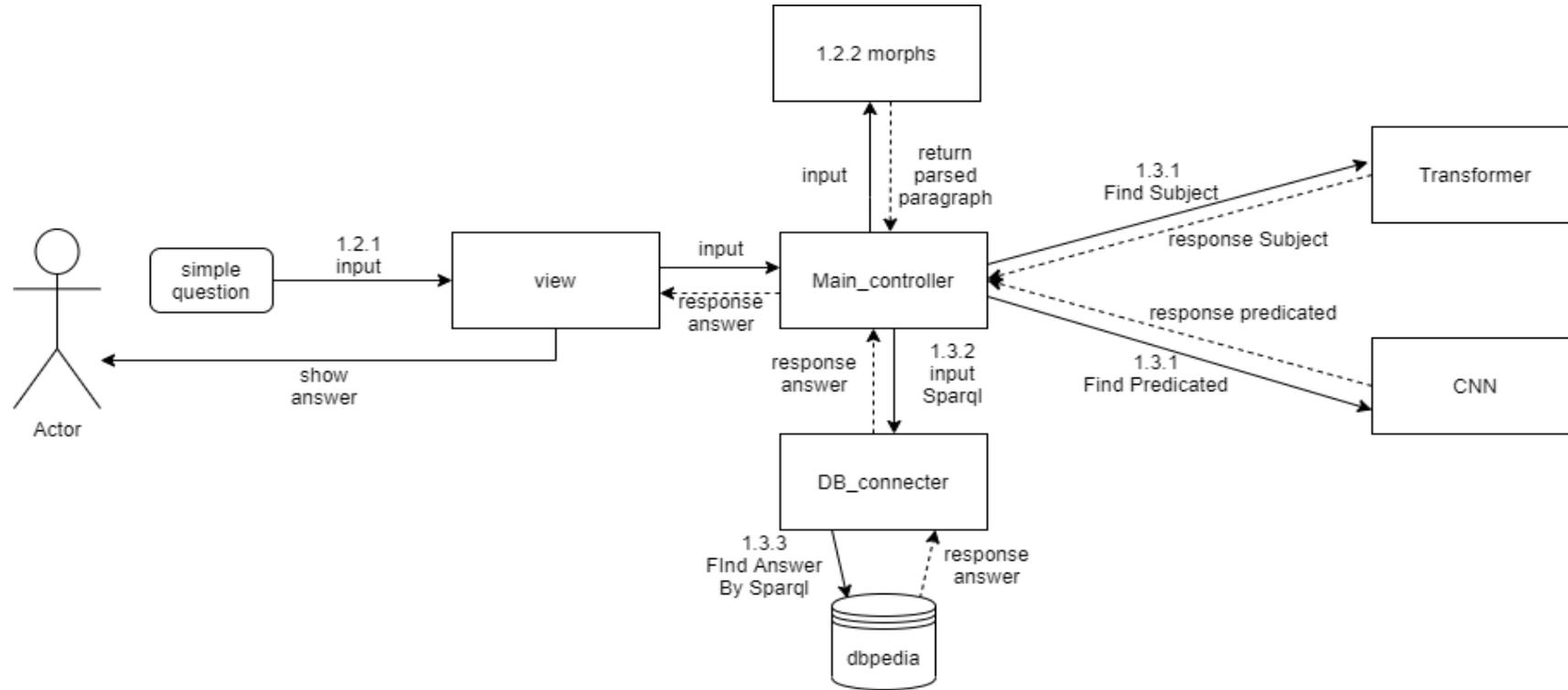
# 002 시스템 테스트 케이스

Case No.	테스트케이스 목표	입력 상황	예상 결과
<b>TEST 1.1.1</b> (기능)	Question 에 대한 Subject 를 정확하게 뽑아낼 수 있는 AI 가 구축되어야 한다.	아메리카노의 가격은 얼마인가요?	“아메리카노”라는 데이터가 도출됨
<b>TEST 1.1.2</b> (기능)	Question 에 대한 predicated 를 정확하게 뽑아낼 수 있는 AI 가 구축되어야 한다.	아메리카노의 가격은 얼마인가요?	“가격”이라는 데이터가 도출됨
<b>TEST 1.2.1</b> (기능)	사용자가 등록한 Simple question 프로그램이 정확하게 입력되어야 한다.	사용자 입력: 아메리카노의 가격은 얼마인가요?	“아메리카노의 가격은 얼마인가요?”라는 문장을 입력받음
<b>TEST 1.2.2</b> (기능)	입력받은 문장에 형태소 분석이 튜플로 제대로 분석이 되어야한다.	아메리카노의 가격은 얼마인가요?	(아메리카노/의/ 가격/은/얼마/ 인가/요/?)

<b>TEST 1.3.1</b> (기능)	입력된 데이터들을 통해서 정확하게 Subject 와 Predicated 를 추출해야 한다.	<b>(아메리카노/의/ 가격/은/얼마/ 인가/요/?)</b>	Subject – 아메리카노 Predicated - 가격
<b>TEST 1.3.2</b> (기능)	입력된 데이터들을 통해서 정확하게 Sparql 가 만들어져야 한다.	<b>Subject – 아메리카노 Predicated - 가격</b>	Select 가격 Where 아메리카노
<b>TEST 1.3.3</b> (기능)	DBPedia 를 참조하여 답변이 정확하게 나와야 한다.	Select 가격 Where 아메리카노	4000
<b>Test 2</b> (비기능)	Simple Question 을 집어넣어 정답률 70% 이상을 달성해야한다.	Simple Question 의 Test Dataset 을 모델에 입력	테스트 결과 70%이상



# 003 상위 디자인 아키텍처 설계



**View** - UI적 요소를 처리하기 위한 모듈(input이나 exposure)

**Main\_controller**- 모듈의 데이터를 총괄하는 모듈

**Morphs** - 입력된 데이터를 형태소 분석하는 모듈

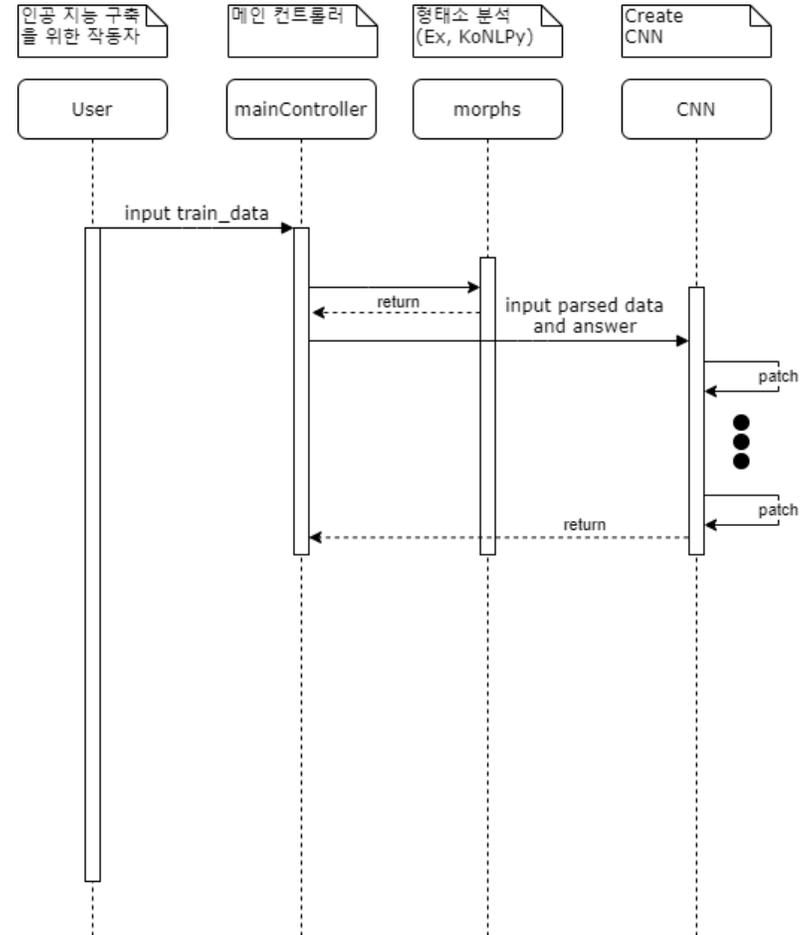
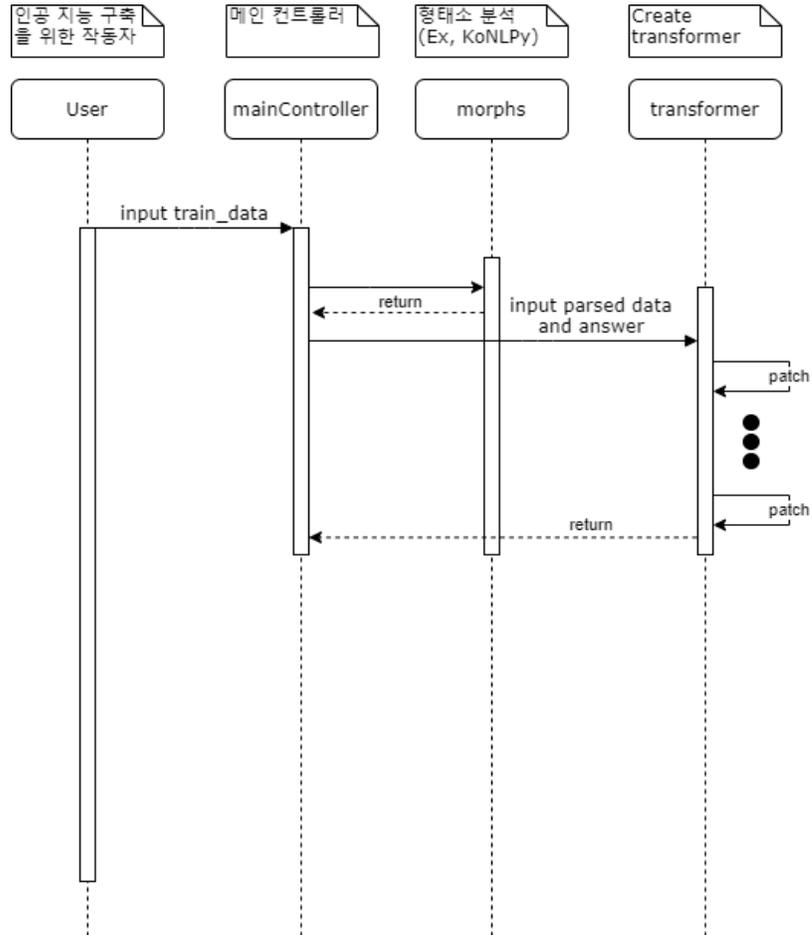
**Transformer** - Subject를 찾기 위하여 생성된 AI 모듈

**CNN** - Predicated를 찾기 위하여 생성된 AI 모듈

**DB\_Connector** - SQL과 DB를 연결하기 위한 모듈

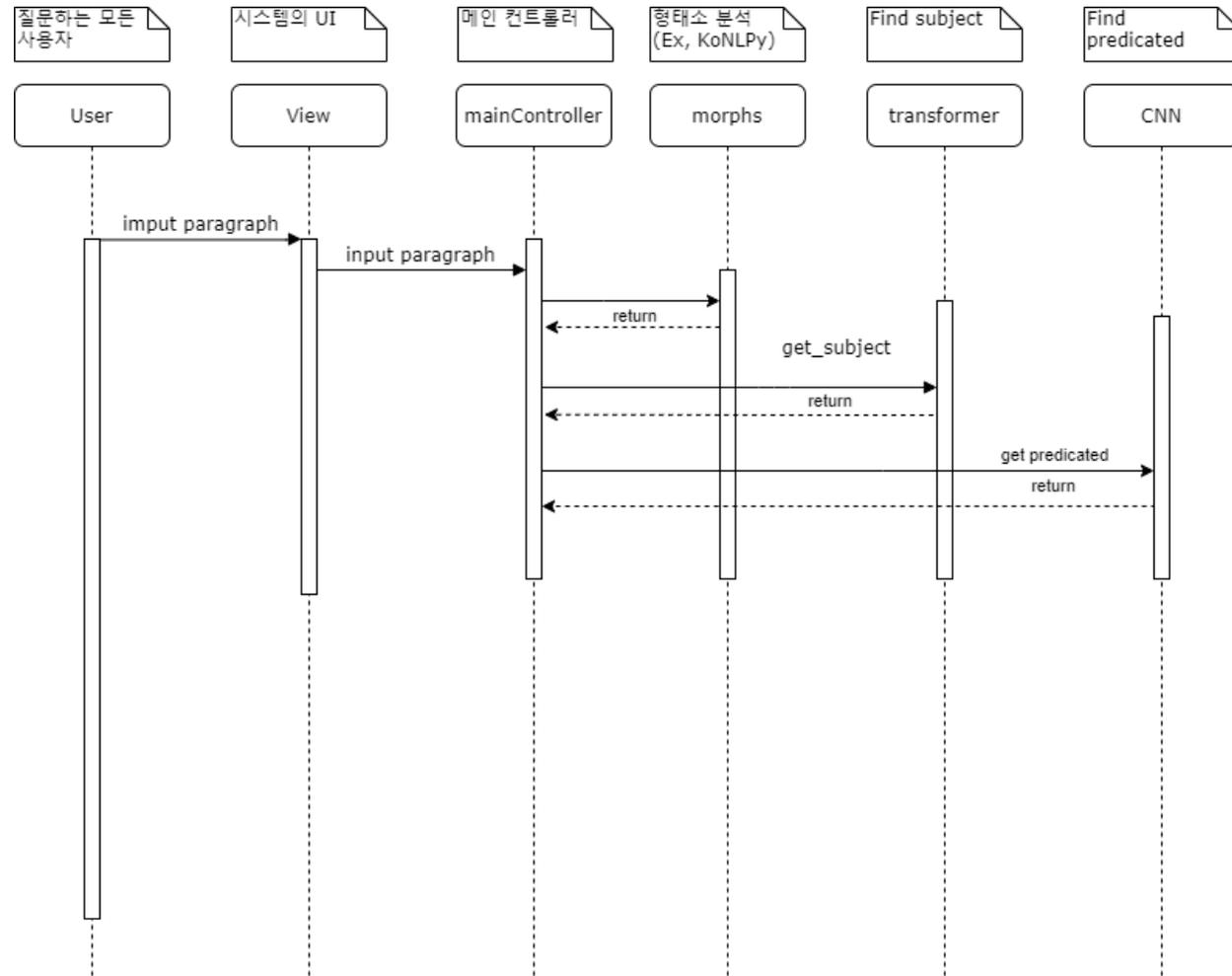


# 003 상위 디자인 시퀀스 다이어그램



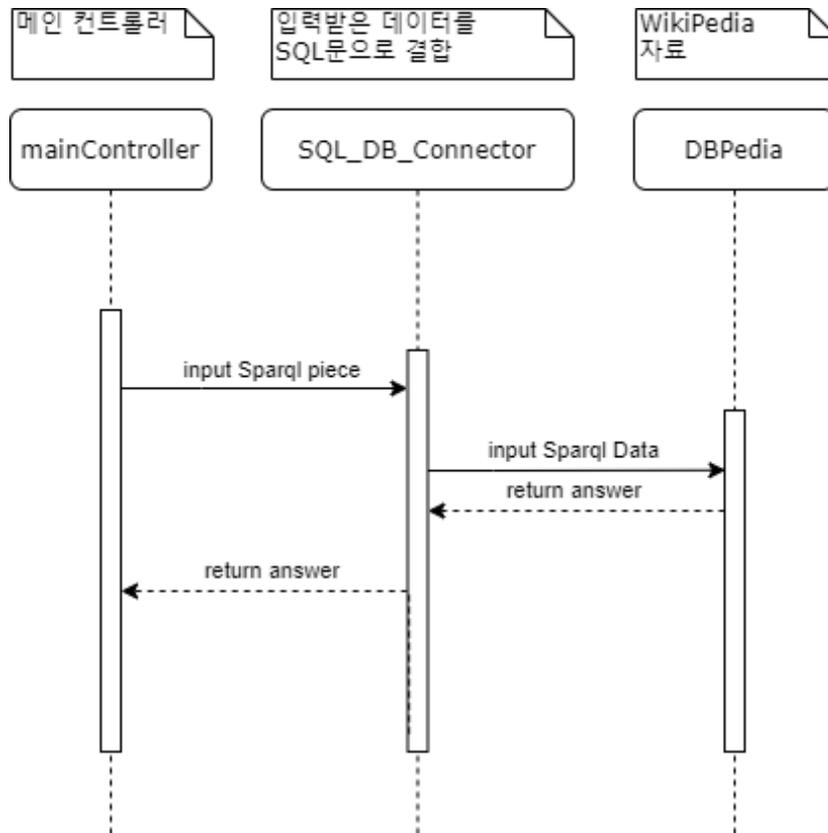
# 003 상위 디자인

## 시퀀스 다이어그램



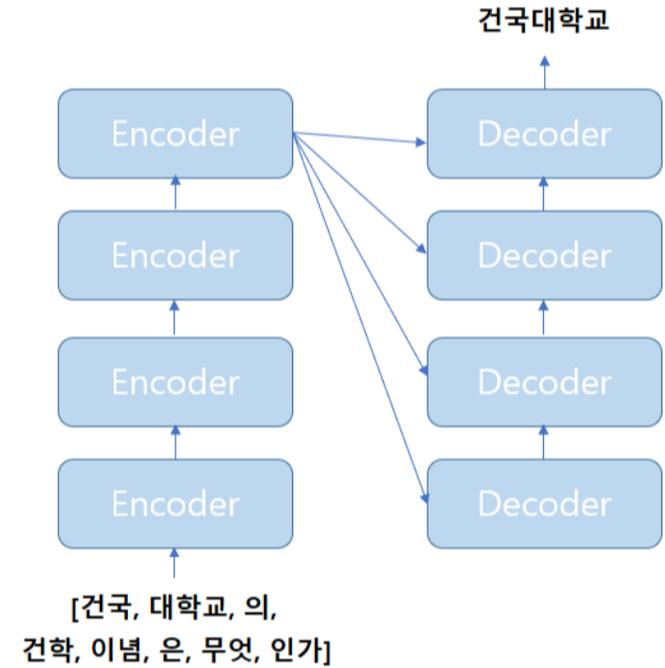
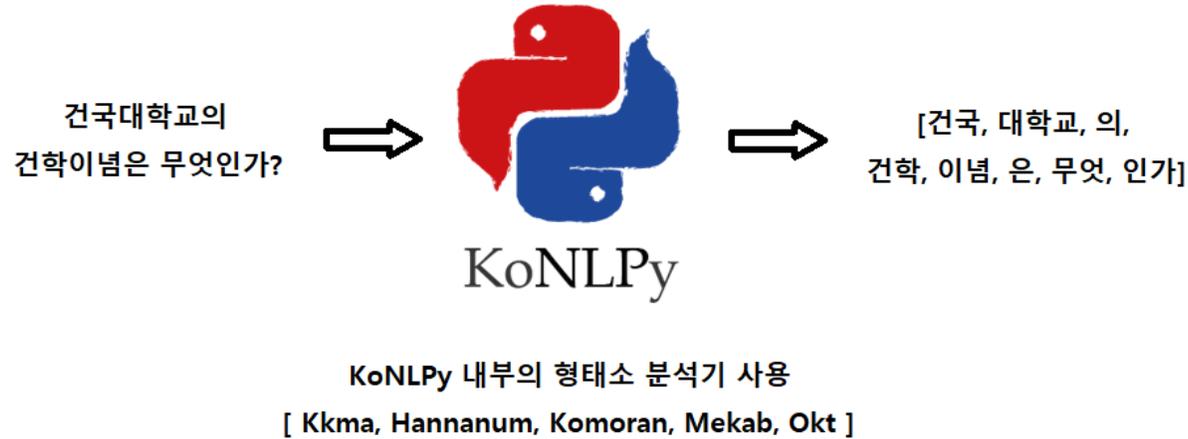
# 003 상위 디자인

## 시퀀스 다이어그램



# 004 상세 디자인

## Morph, Transformer



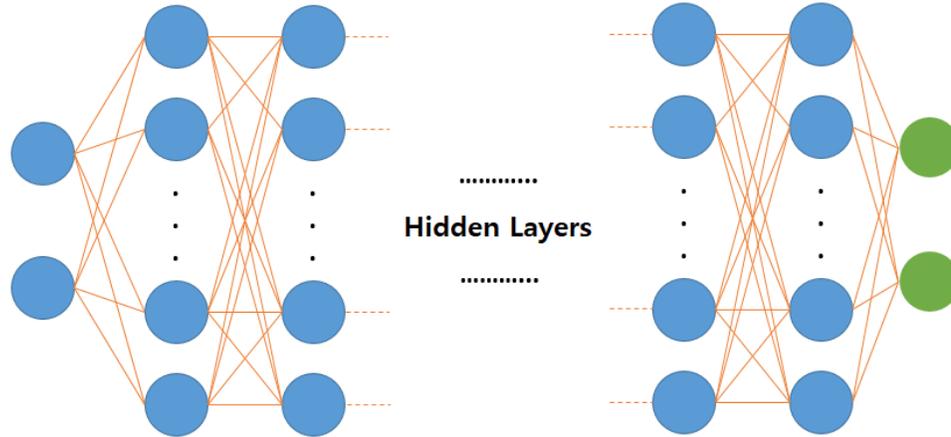
<< TRANSFORMER >>



# 004 상세 디자인

## CNN

NL문장에서 올바른 RELATION을 도출하도록 AI모델을 학습시킨다



**Input Nodes :**  
형태소 분석 결과  
[건국, 대학교, 의, 건학, 이념, 은,  
무엇, 인가, Null, Null, Null .... ]

**OutPut Nodes:**  
DBPedia의 모든 테이블의  
칼럼명의 합집합  
약 2만 개



# 005 Traceability Matrix

