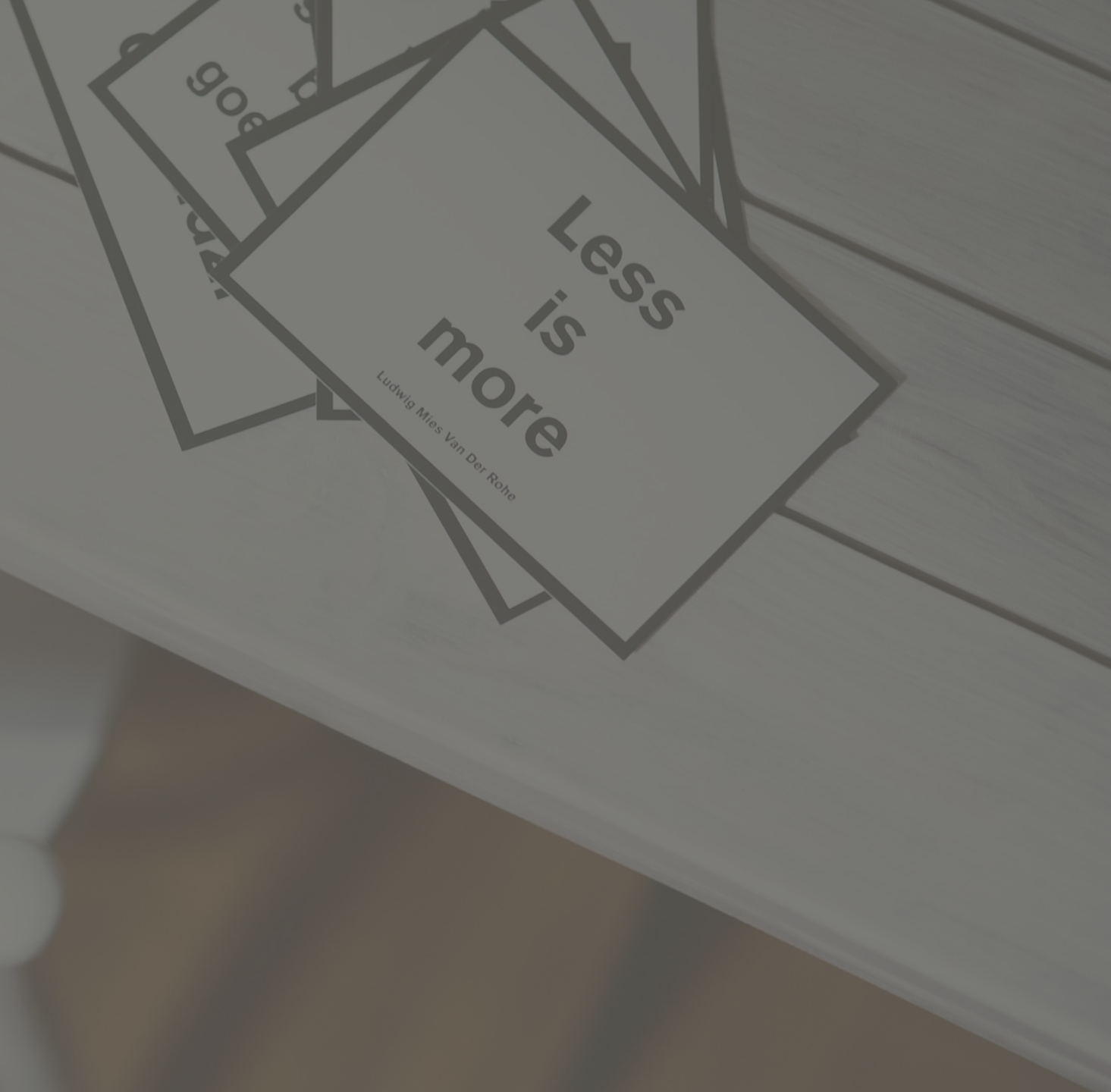


2020 졸업프로젝트 :

한국어 Simple Question의 SPARQL 변환

- 001 프로젝트 소개
- 002 최종 산출물
- 003 1차 구현물
- 004 Pass/Fail Criteria
- 005 System Test Result
- 006 Component Diagram
- 007 2nd Iteration 계획



001 프로그램 소개

- NLP to sparql 연구는 많음
 - But! 한국어 NLP to sparql 자료는 찾기가 매우 힘들
- 한국어 NLP to Sparql은 영어 보다 어려움
 - 영어는 띄어쓰기로 구분하지만 한글은 품사 구분도 필요
- 한국어 dbpedia를 참조하여 NLP to sparql을 구현을 목표

002 최종 산출물

지식 기반 한국어 QA 프로그램

질문을 입력해 주세요.

Find Answer

출력된 값입니다.

지식 기반 한국어 QA 프로그램

야나미 조지가 태어난 곳은 어디인가?

Find Answer

도쿄시

003 1차 구현물

- 한국어 디비피디아 16년도 dump를 사용
- 디비피디아 트리플 데이터를 이용해 simple question dataset 생성
- Subject, relation 추출 성공 가정 하에 sparql 쿼리 생성 및 데이터 접근 성공
- 형태소 분석을 통한 토큰화 작업

```
>>> gen.query('아이작_뉴턴', 'field')
```

```
['경제학', '기독교_신학', '수학', '물리학', '자연철학', '천문학', '연금술']
```

	Subject	Relation	Object	Question
0	죽음의_트라이앵글	starring	이오아나_몰도반	죽음의_트라이앵글에 출연한 배우는 누구인가?
22	TAKE_FIVE	starring	카라사와_토시아키	TAKE_FIVE에 출연한 배우는 누구인가?
31	에브리바디_올라잇	starring	마크_러펄로	에브리바디_올라잇에 출연한 배우는 누구인가?
32	나는_가수다_3	starring	김연우	나는_가수다_3에 출연한 배우는 누구인가?
50	사랑은_노래를_타고	starring	곽희성	사랑은_노래를_타고에 출연한 배우는 누구인가?

551972	파더_앤_도터	starring	에런_폴	파더_앤_도터에 출연한 배우는 누구인가?
551980	13세_소년	starring	김정훈	13세_소년에 출연한 배우는 누구인가?

```
↳ [('야', 'NNG'),
    ('나미', 'NNG'),
    ('조지', 'NNG'),
    ('가', 'JKS'),
    ('태어나', 'VV'),
    ('니', 'ETD'),
    ('곳', 'NNG'),
    ('은', 'JX'),
    ('어디', 'NP'),
    ('이', 'VCP'),
    ('니가', 'EFQ'),
    ('?', 'SF')]
```

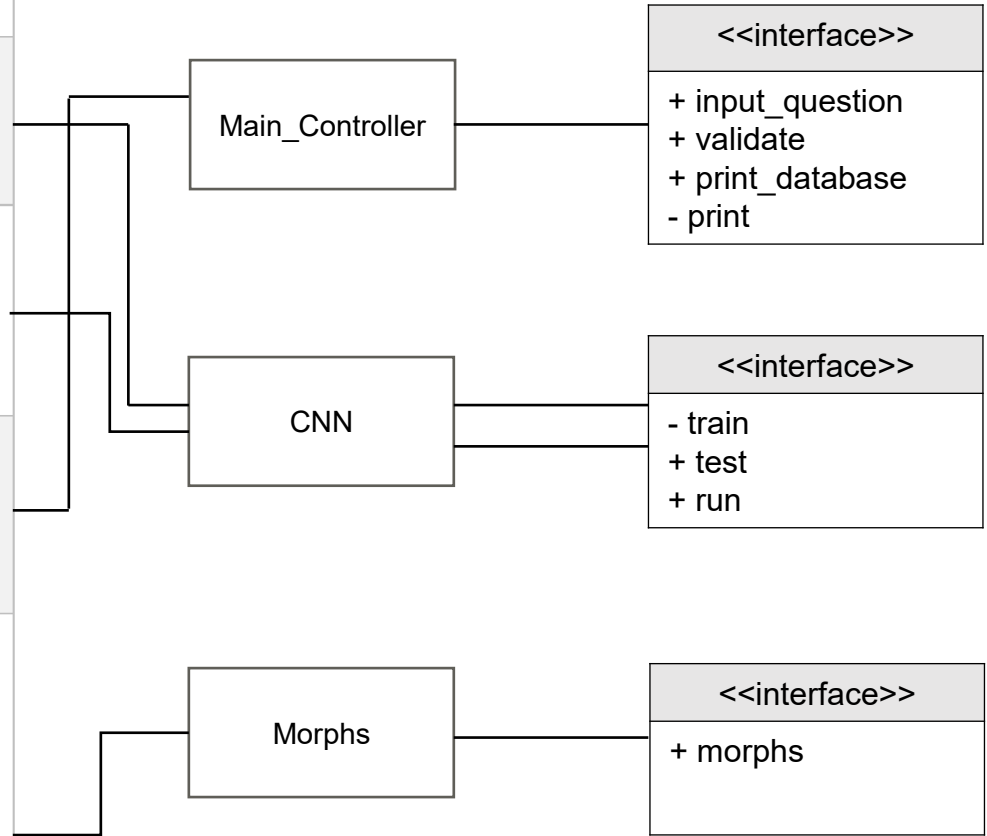
004 Pass/Fail Criteria

Simple Question 정확도 70% 달성

- 형태소 분석을 Kkma를 써보며 정답률을 조정하기 위해서 다른 형태소 분석기 역시 사용 (Kkma(꼬꼬마)의 형태소 분석기 결과가 81%로 가장 높은 결과이다)
- NLP to SQL 정확도 70% 달성을 목표. (결과의 기준: Question을 집어넣었을 경우 subject와 relation이 정확하게 이끌어 낼 수 있는지를 확인).
- 위의 결과와 별개로 object 결과가 올바른 것을 확인하기 위하여 DBPedia 데이터를 이용한 질의응답 정답률도 확인. 이 역시 70% 달성을 목표
- 성공 여부의 기준을 변경하여 최적의 방법 탐색

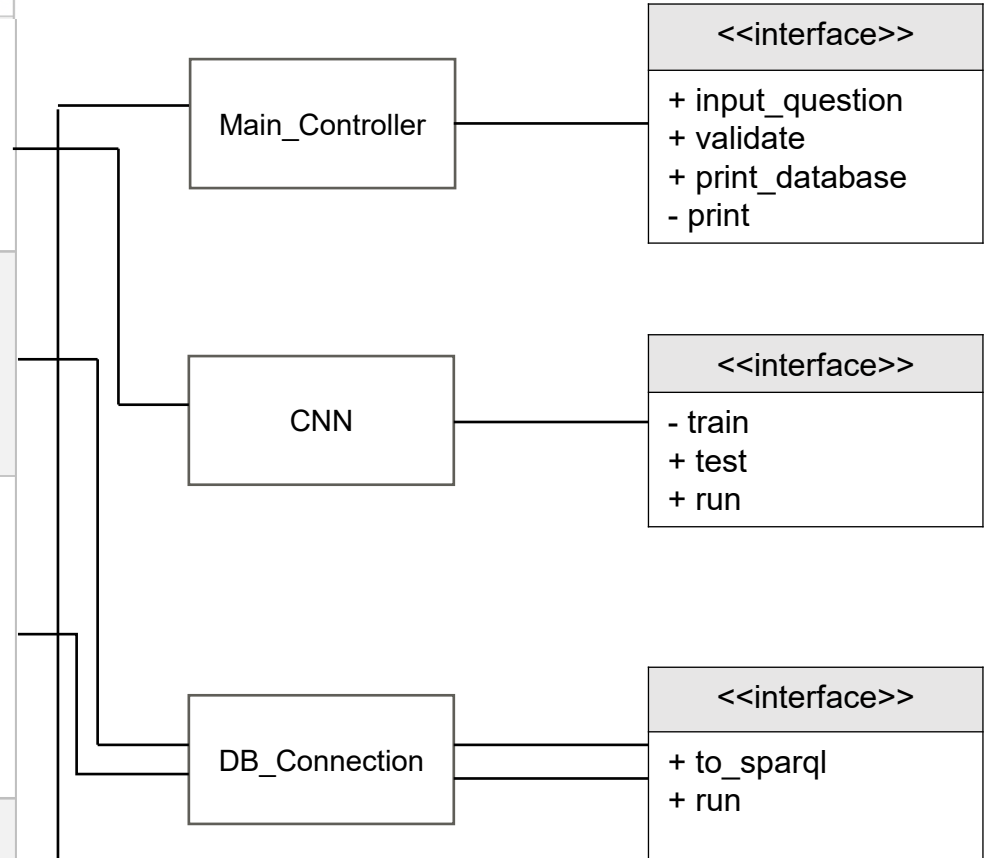
005 System Test Result

Case No.	테스트케이스 목표	입력 상황	예상 결과	실행 결과
TEST 1.1.1 (기능)	Question 에 대한 Subject 를 정확하게 뽑아낼 수 있는 AI 가 구축되어야 한다.	야나미 조지가 태어난 곳은 어디인가?	“도쿄시”라는 데이터가 도출됨	실패 - AI 가 구축되지 않았다.
Test 1.1.2 (기능)	Question 에 대한 predicated 를 정확하게 뽑아낼 수 있는 AI 가 구축되어야 한다.	야나미 조지가 태어난 곳은 어디인가?	“태어난 곳 (birthPlace)”이라는 데이터가 도출됨	실패 - AI 가 구축되지 않았다.
TEST 1.2.1 (기능)	사용자가 등록한 Simple question 프로그램이 정확하게 입력되어야 한다.	사용자 입력: 야나미 조지가 태어난 곳은 어디인가?	“야나미 조지가 태어난 곳은 어디인가?”라는 문장을 입력받음	성공 - ‘야나미 조지가 태어난 곳은 어디인가?’ 가 인풋 문장으로 입력된다.
TEST 1.2.2 (기능)	입력 받은 문장에 형태소 분석이 튜플로 제대로 분석이 되어야한다.	야나미 조지가 태어난 곳은 어디인가?	(야/나미/조지/가/ 태어나/ㄴ/곳/은/ 어디/ 이/ㄴ가/?/)	성공 - [(‘야’, ‘NNG’), (‘나미’, ‘NNG’), (‘조지’, ‘NNG’), (‘가’, ‘JKS’), (‘태어나’, ‘VV’), (‘ㄴ’, ‘ETD’), (‘곳’, ‘NNG’), (‘은’, ‘JX’), (‘어디’, ‘NP’), (‘이’, ‘VCP’), (‘ㄴ가’, ‘EFQ’), (‘?’ , ‘SF’)]

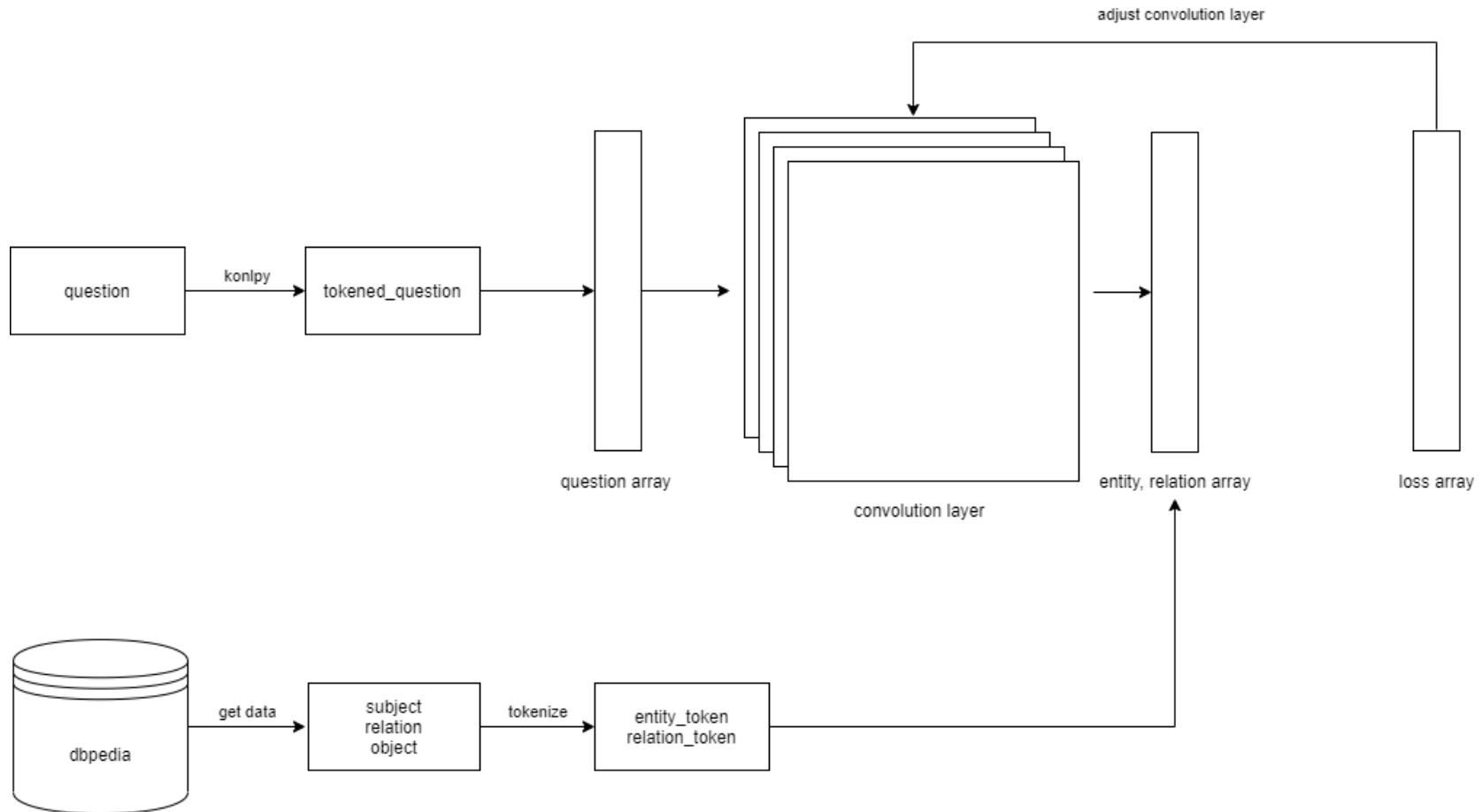


005 System Test Result

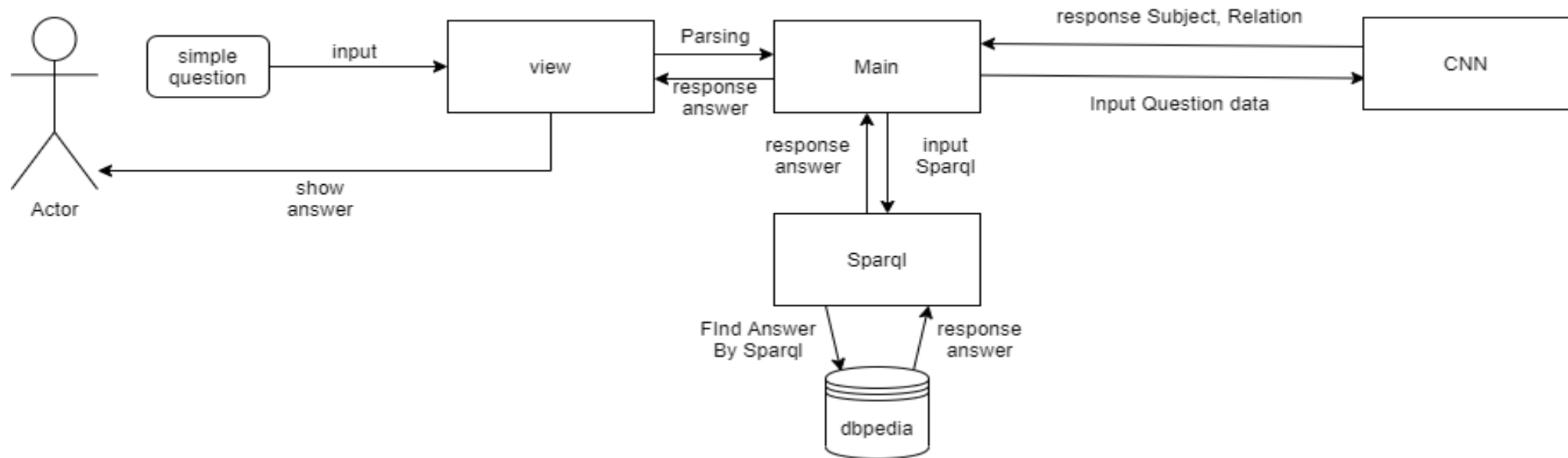
Case No.	테스트케이스 목표	입력 상황	예상 결과	실행 결과
TEST 1.3.1 (기능)	입력된 데이터들을 통해서 정확하게 Subject 와 Predicated 를 추출해야 한다.	야나미 조지가 태어난 곳은 어디인가?	Subject - 야나미 조지 Predicated - 태어난 곳	실패 - AI 가 구축되지 않았다.
TEST 1.3.2 (기능)	입력된 데이터들을 통해서 정확하게 Sparql 가 만들어져야 한다.	Subject - 야나미 조지 Predicated - 태어난 곳	Select 태어난 곳 Where 야나미 조지	성공 - SELECT ?b WHERE { <http://ko.dbpedia.org/resource/야나미_조지> <http://dbpedia.org/ontology/birthPlace> ?b . }
TEST 1.3.3 (기능)	DBPedia 를 참조하여 답변이 정확하게 나와야 한다.	SELECT ?b WHERE { <http://ko.dbpedia.org/resource/야나미_조지> <http://dbpedia.org/ontology/birthPlace> ?b . }	도쿄시	성공 -'도쿄시' 결과가 나온다.
Test 2 (비기능)	Simple Question 을 집어넣어 정답률 70% 이상을 달성해야한다.	Simple Question 의 Test Dataset 을 모델에 입력	테스트 결과 70%이상	실패 - 테스트를 진행할 만큼 프로그램 구축이 진행되지 않음



006 Component Diagram



006 Component Diagram



View - UI적 요소를 처리하기 위한 모듈(input이나 exposure)

Main_controller- 모듈의 데이터를 총괄하는 모듈

Morphs - 입력된 데이터를 형태소 분석하는 모듈

CNN - Predicated를 찾기 위하여 생성된 AI 모듈

DB_Connecter - SQL과 DB를 연결하기 위한 모듈

007 2nd Iteration 계획

1. CNN 구축

2. 정답률 개선 노력

- 형태소 분석기 변경
- 데이터 입력 방법 변경 (띄어쓰기 고려 등)
- CNN 레이어 개수/크기 변경

3. 질문 데이터 생성

- train data
- test data

5/28까지 CNN 구축

- Natural language로부터 subject와 relation 추출
- zoom과 미팅을 통한 정기적인 회의