

중간 발표 2차

For

시각 장애인을 위한 상황 묘사 어플리케이션

Team 4

정준호(201113275) 전민규(201411802) 김도연(201614157)



Contents



프로젝트 소개

1st Iteration

2nd Iteration

System Testing Results

Pass / Fail Criteria

Component Diagram

3rd Iteration 계획

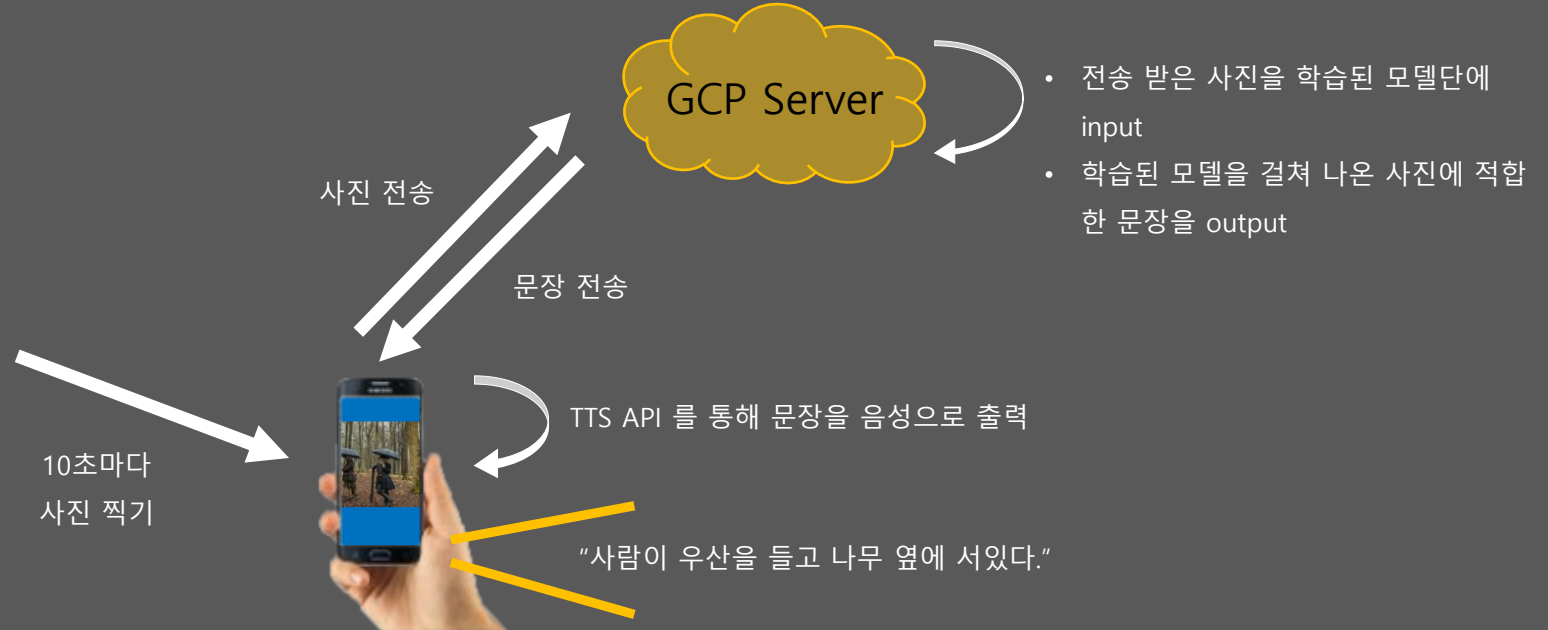
시각 장애인을 위한 상황 묘사 어플리케이션

' 시각 장애인을 위한 상황 묘사 어플리케이션 ' 이란 ?

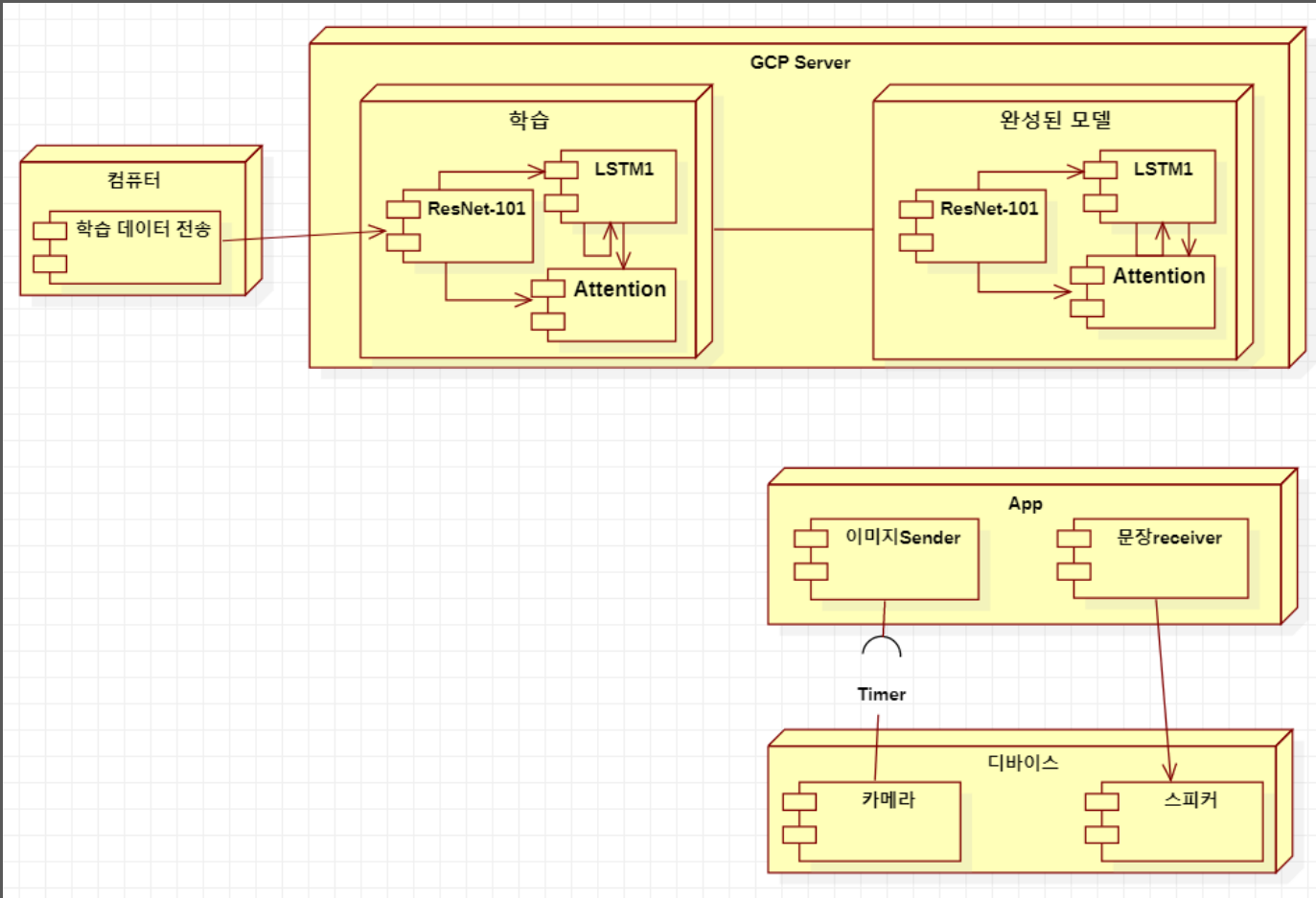
시각 장애인분들을 위해 접근성이 쉬운 핸드폰 카메라로 찍히는 사진의 상황을 글 및 소리로 묘사해주는 어플리케이션입니다.

' 시각 장애인을 위한 상황 묘사 어플리케이션 ' 이 작동하는 방식은 ?

스마트폰을 들고 있으면 자동적으로 10초마다 한 장씩 사진을 찍어 그에 대한 설명을 음성으로 들려줍니다.



1st Iteration



APP과 서버가 따로 구현

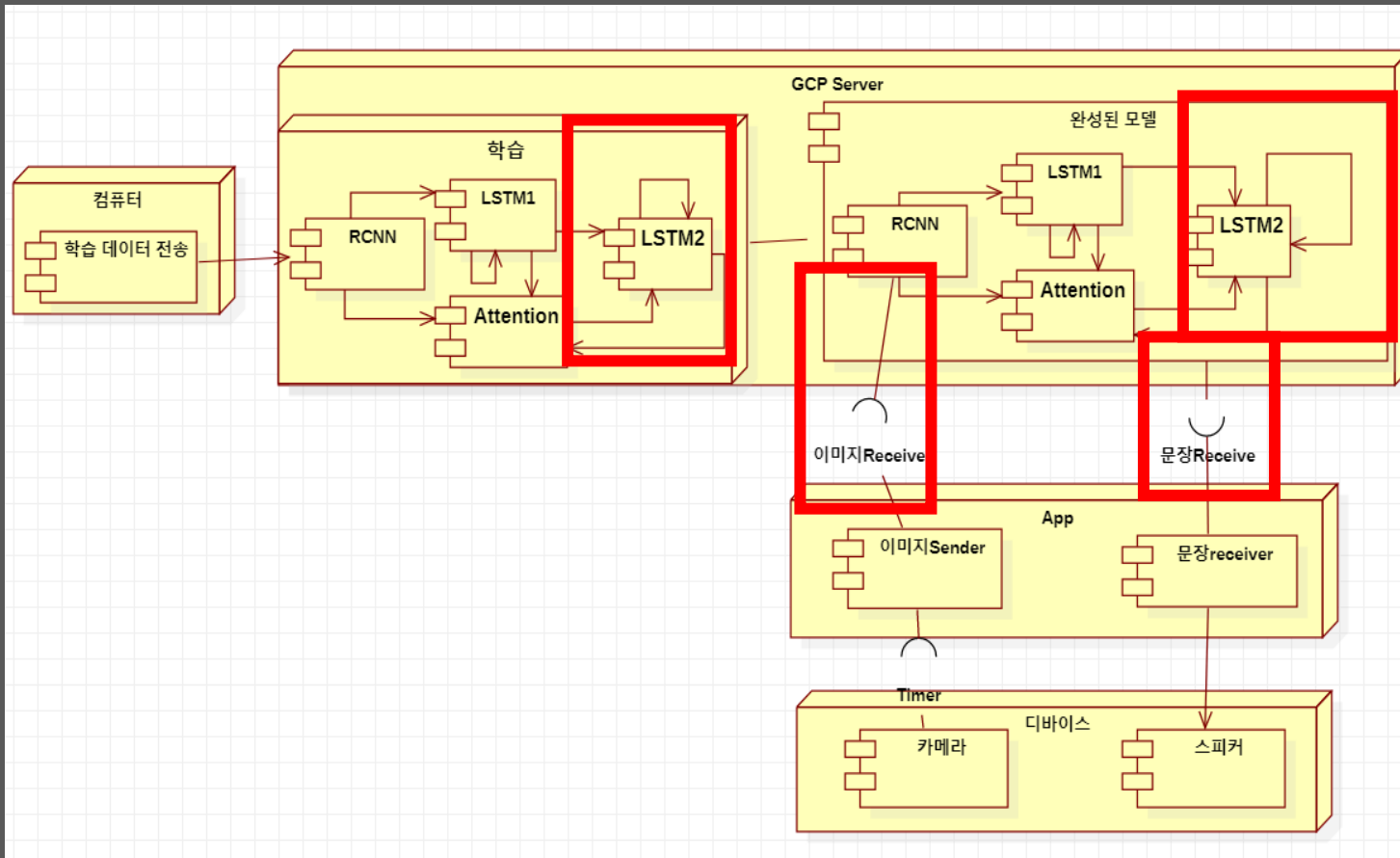
< APP >

- 디바이스에게 사진을 받음
- TTS API를 통해 문장을 음성으로 변환하여 디바이스에게 전달

< GCP Server >

- 서버안에서 MSCOCO 이미지데이터를 input시켜 학습
- 학습시킨 모델을 통해 이미지를 전달받으면 이미지에 알맞은 문장을 출력

2nd Iteration



APP과 서버를 연결시킴
성능을 위해 LSTM2 추가

< APP >

- 서버와 연결

< GCP Server >

- 학습시키는 모델과 완성된 모델에 LSTM2을 추가

2차 구현 (데모 영상)



System Testing Results

TEST CASE	NAME	DESCRIPTION	Pass / Fail
1-1	전면부 상황 입력	<ul style="list-style-type: none">스마트폰 앱에서 10초마다 사진이 찍히는지 테스트 한다.	Pass
1-2	전면부 상황 입력	<ul style="list-style-type: none">새로 핸드폰 카메라로 '사람이 우산을 들고 나무 옆에 있는 사진'을 찍고 그 사진을 이미 GCP에서 학습된 weight를 갖고 있는 모델의 입력으로 넣고, 모델 입력 단에서 카메라로 찍은 사진 frame이 들어오는지 확인 하는 코드를 추가한다.Input image가 잘 들어오면 'OK', 아니면 'No Image'라는 메시지를 코드단에서 출력하도록 한다.	Pass
2	딥러닝 학습을 통한 모델	<ul style="list-style-type: none">Test data(train에서 사용하지 않은 MSCOCO 8000장)로 모델을 평가하였을 때, 평균적인 BLUE Score-4 기준 어절단위 0.111 의미형태소 단위 0.225 형태소 단위 0.251 이상이면 신뢰도를 갖고 있다고 판단한다.	Fail
3	학습된 모델을 통한 결과 도출	<ul style="list-style-type: none">위에서 정의한 '사람이 우산을 들고 나무 옆에 있는 사진' 과 그에 맞는 문장 '사람이 우산을 들고 나무 옆에 서있음' 을 준비해두고, 준비해둔 사진을 모델의 입력으로 넣었을 때, 미리 준비해둔 그 사진을 설명하는 문장 '사람이 우산을 들고 나무 옆에 서있음'과 유사한 문장이 10초 이내에 출력 되는지 확인한다.	Pass
4	TTS (Text To Speech)	<ul style="list-style-type: none">결과로 나온 문장을 TTS API를 통해 음성으로 바꿔서 출력하는지를 체크한다.	Pass

Pass / Fail Criteria

TEST CASE	NAME	Pass / Fail	Reason
1-1	전면부 상황 입력	Pass	→ 10초마다 정상적으로 사진이 찍힌다.
1-2	전면부 상황 입력	Pass	→ 2nd Iteration에서 앱과 서버를 연결시킴으로 앱이 전달한 사진이 완성된 모델단에 오류없이 input된다.
2	딥러닝 학습을 통한 모델	Fail	<ul style="list-style-type: none"> → 모델은 Encoder/Attention/Decoder 다 구현하였다. Ex) Encoder : RCNN // Decoder : LSTM1 & LSTM2 구현 → LSTM2를 추가하고 학습시키려고 보니 2nd Iteration까지 시간이 얼마 남지않아 학습을 충분하게 하지 못하였다. → 일단 train 시켜본 결과, 정했던 BLUE Score-4 기준에 도달하지 못했다. → 또한 실제 사진을 input하여 나온 결과를 확인하여 보니 사진의 상황과 부합하지 못하는 문장을 만들어내는 것을 볼 수 있었다.
3	학습된 모델을 통한 결과 도출	Pass	→ 앱과 서버를 연결시켜 앱이 디바이스의 카메라에게 사진을 요청한 순간부터 문장을 TTS API를 통해 음성으로 바꾼 뒤 디바이스의 스피커로 나오는 시간까지 10초안에 이루어지는 것을 확인할 수 있었다.
4	TTS (Text To Speech)	Pass	→ 앱에서 그냥 임의의 문장을 출력하게 했을 때, 문장을 정확하게 읽었다.

Final Iteration 계획

TEST CASE	NAME	Pass / Fail	Plan
1-1	전면부 상황 입력	Pass	X
1-2	전면부 상황 입력	Pass	X
2	딤러닝 학습을 통한 모델	Fail	Final Iteration 까지 충분하게 학습을 시켜 사진의 상황에 알맞은 문장으로 출력하도록 하겠습니다.
3	학습된 모델을 통한 결과 도출	Pass	X
4	TTS (Text To Speech)	Pass	X

Q & A

감사합니다.