

# 챗봇을 활용한 VQA 서비스 플랫폼

Euclid Soft

장주영   류세영   김영희



장주영(25)

팀장

Json 분석

기능 구현

시스템 흐름도 작성



류세영(24)

팀원

오류검출

보완 및 기능 수정



김영희(23)

팀원

요구사항 분석

가이드 작성



## 프로젝트 일정

Work Break-down Structure									11월				
									2주				
Procedures	Steps	Tasks	담당자	schedule		산출물/비고	담당	월	화	수	목	금	
				시작일	종료일			2	3	4	5	6	
1.0 분석/설계													
	1.1	분석	요구사항분석										
	1.2	설계	스토리보드 작성		2020.10.26	2020.10.27	스토리보드	공통					
			기능 리스트 작성		2020.10.27	2020.10.27	체크리스트	공통					
			ERD 작성		2020.10.28	2020.10.28	ERD	공통					
2.0 텔레그램 Chatbot을 활용한 VQA 데이터 축적 서비스 개발													
대분류	중분류	소분류	세분류	세분류체크항목									
Django	기능	프로젝트 생성	환경설정	DB 연동	2020.10.30	2020.11.02	공통						
		Telegram	Chatbot 생성		2020.10.29	2020.10.29	공통						
			JSON 분석	Response 메시지	2020.11.03	2020.11.03	장주영						
			기능	사용방법	2020.11.03	2020.11.04	류세영						
				시작	2020.11.04	2020.11.04	류세영						
				이미지전송	2020.11.03	2020.11.04	장주영						
				질문	2020.11.03	2020.11.03	김영희						
				답변	2020.11.04	2020.11.04	김영희						
				공유	2020.11.05	2020.11.05	공통						
		보완	username 없을경우	2020.11.06	2020.11.09	류세영							
3.0 테스트 및 완료													







# CONTENTS

01

---

개 요

02

---

개발방법

03

---

사용방법

04

---

결 론



# 01

---

개 요

## 동기 및 필요성

광주시가 인공지능 시장의 밑거름이 되는 데이터 산업을 활성화하고 일자리를 창출하기 위해 (주)유클리드소프트와 함께 광주시민을 대상으로 인공지능 데이터 라벨링 교육을 한다.

< '광주시, 인공지능 데이터 라벨링 교육생 2차 모집', 천지일보이미에 기자 ([lme381@newsci.co.kr](mailto:lme381@newsci.co.kr)) >

일자리에 참여하는 청년들은 생활 환경 4종의 이미지 데이터 100만장과 텍스트 Q&A 700만 쌍을 구축하게 되며 특히 이번 사업은 디지털로 수행되는 만큼 시간과 장소에 구애 받지 않는다는 것이 가장 큰 특징이다.

< '청년 디지털 일자리 사업 추진, 박희운 기자 ([hypark@sedaily.com](mailto:hypark@sedaily.com)) >

## 동기 및 필요성



장소에 구애받지 않고  
실시간 데이터 수집이  
가능한 서비스 필요



기존의 복잡한 데이터  
등록방식의 단순화 필요



데이터 입력에 따른  
가이드 및 피드백 제공 필요

# 서비스 설명

Chatbot을 이용한 VQA 데이터 수집

EveryTime, EveryWhere

손쉽게 데이터를 등록받기 위한 서비스 필요

## 목 표

‘ 데이터 제공자로부터 Telegram Chatbot을 통해  
데이터(이미지, 질문, 답)를 제공받고 동시에  
양방향 소통이 가능한 서비스 개발 ’

02

---

개발방법

## 개발 도구

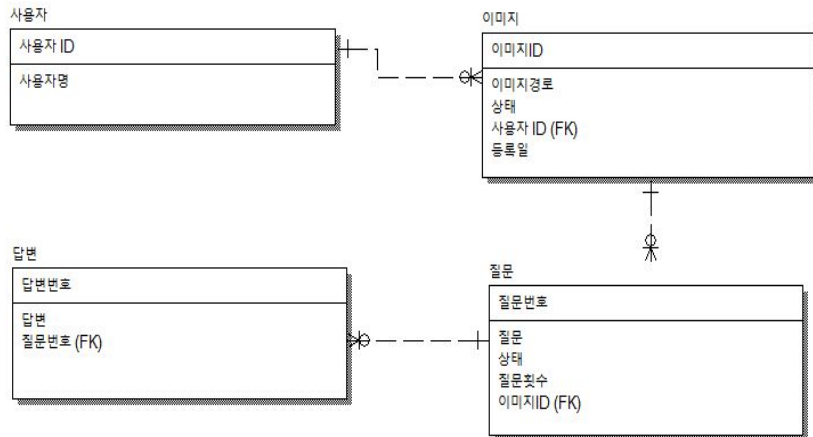


Telegram

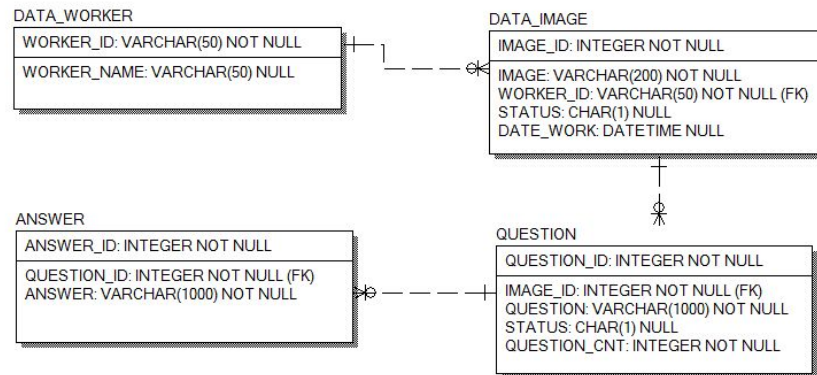
## ERD

Entity Relationship Diagram

## Logical



## Physical





## Request JSON

Text

b'{

```

    "update_id":872364571,Wn
    "message":{
        "message_id":3,
        "from":{
            "id":1344511757,
            "is_bot":false,
            "first_name":"WWuc8fcWWuc601",
            "last_name":"WWuc7a5",
            "username":"hongikchatbot",
            "language_code":"ko"},
        "chat":{
            "id":1344511757,
            "first_name":"WWuc8fcWWuc601",
            "last_name":"WWuc7a5",
            "username":"hongikchatbot",
            "type":"private"
        },
        "date":1604031974,
        "text":":json WWubd84WWuc11d"
    }

```

}'

Image

b'{

```

    "update_id":872364574,Wn
    "message":{
        ...
        "photo":[{
            "file_id":"file_id",
            "file_unique_id":"file_unique_id",
            "file_size":12154,
            "width":320,
            "height":167
        }],
        {
            "file_id":"file_id",
            "file_unique_id":"file_unique_id",
            "file_size":36158,
            "width":690,
            "height":360}
    }

```

}'

# Request JSON

## Image 파일

```
b'{"update_id":872364968,Wn"
  "message":{"
    ...
    "document":{
      "file_name":"20201107_223433.jpg",
      "mime_type":"image/jpeg",
      "file_id":"file_id",
      "file_unique_id":"file_unique_id",
      "file_size":40156,
      "width":320,
      "height":320},
      "file_id":"file_id",
      "file_unique_id":"file_unique_id",
      "file_size":3158609
    }
  }
}
```

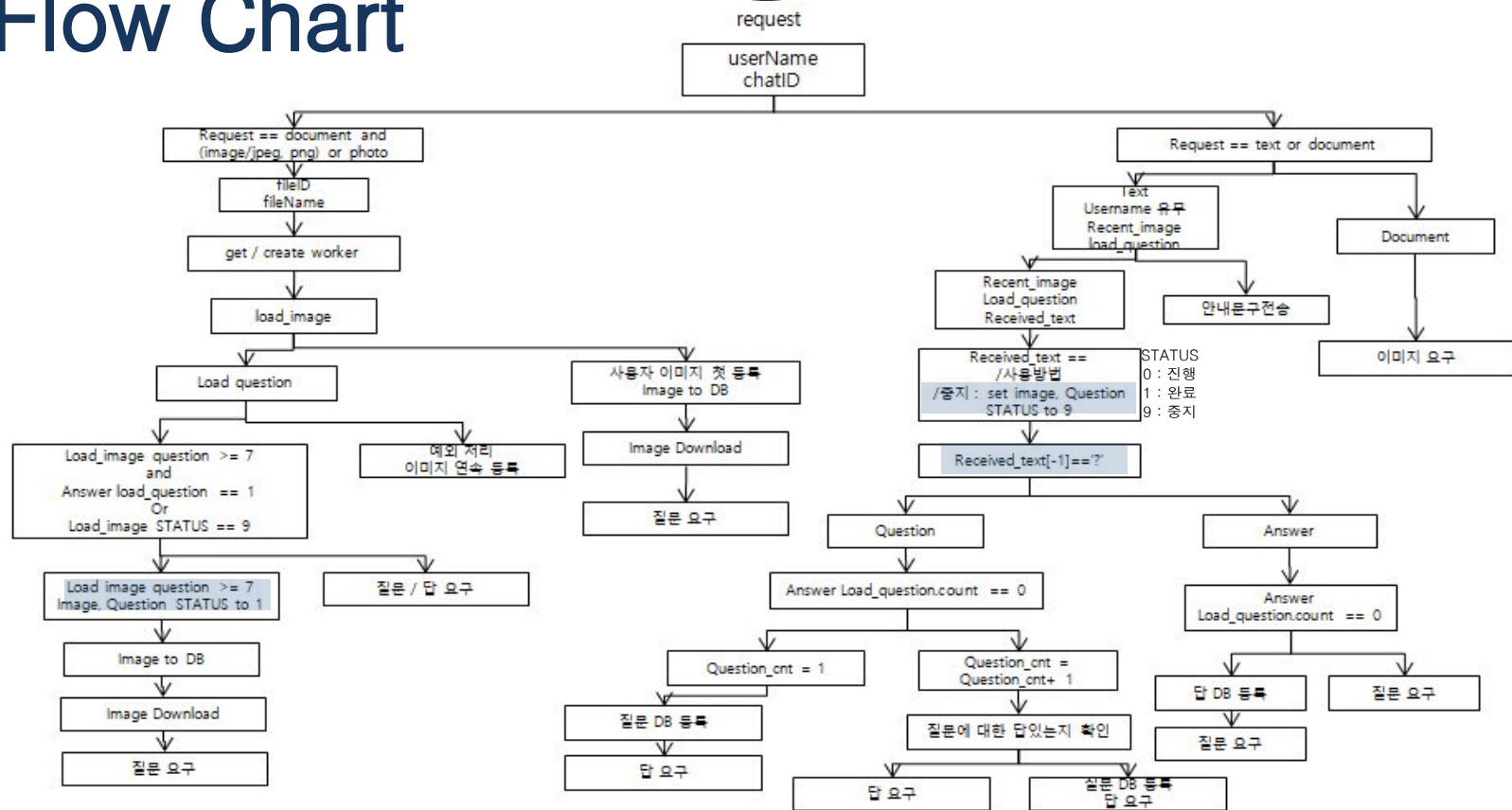
## PDF 파일

```
"file_name":"HW#4-Solution.pdf",
"mime_type":"application/pdf",
```

## Video 파일

```
"file_name":"20201110_202608.mp4",
"mime_type":"video/mp4",
```

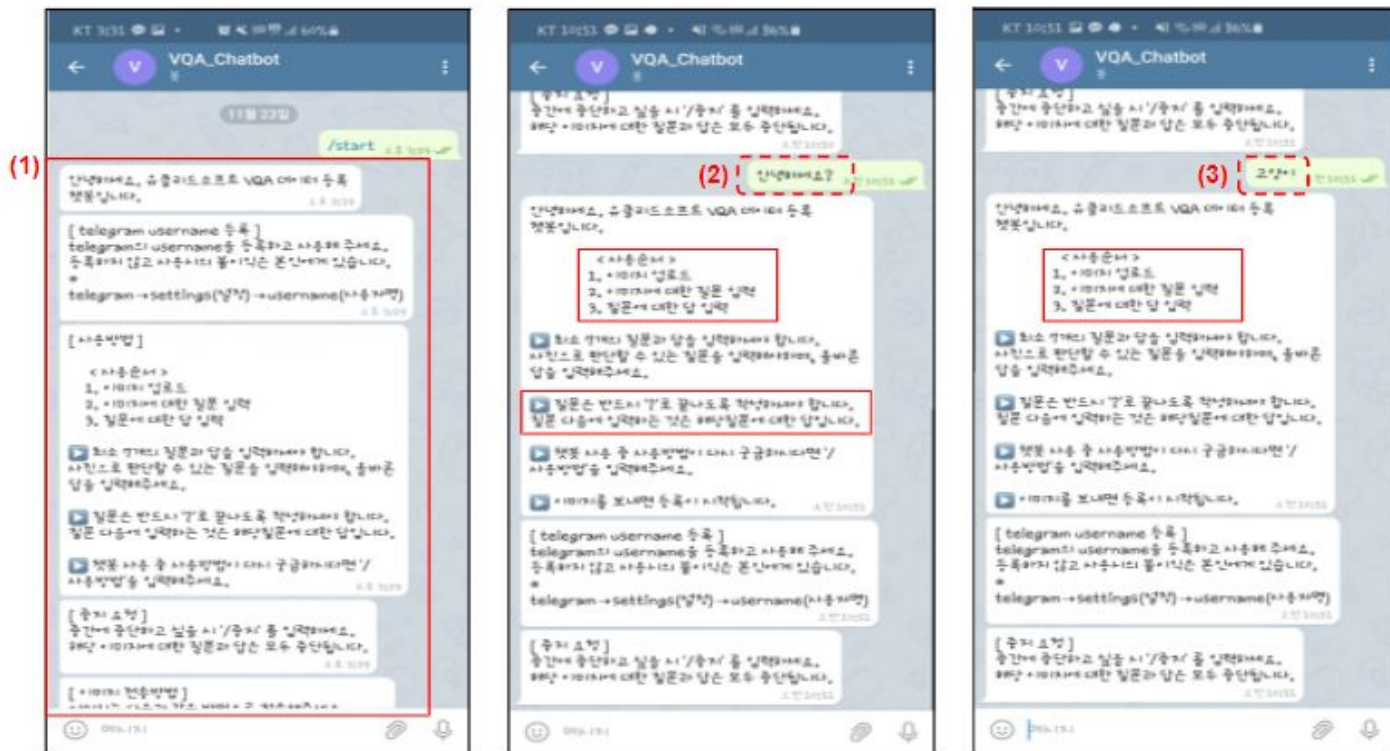
# Flow Chart

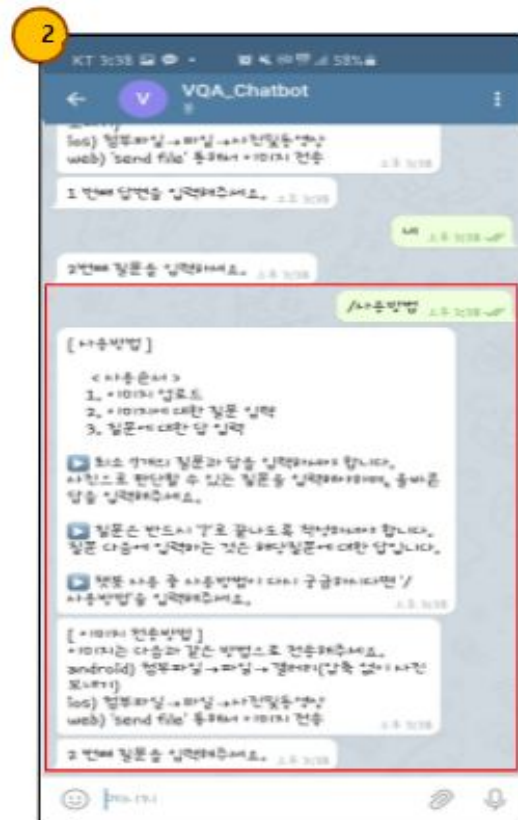
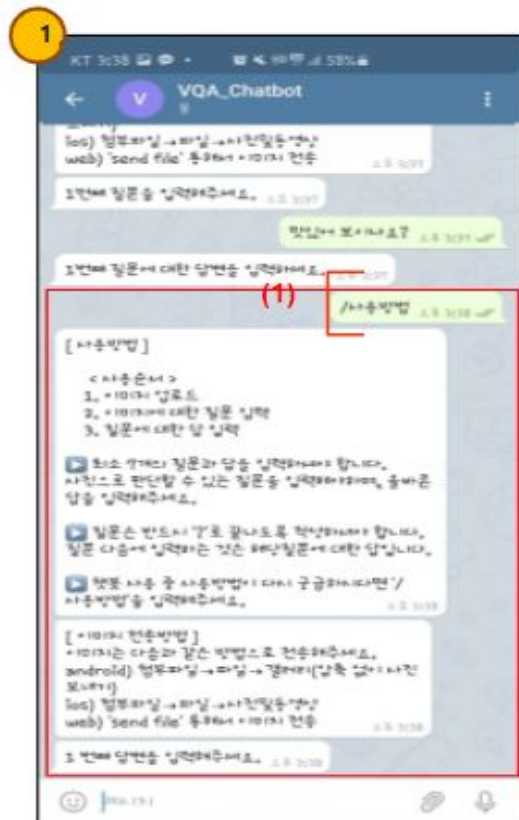


03

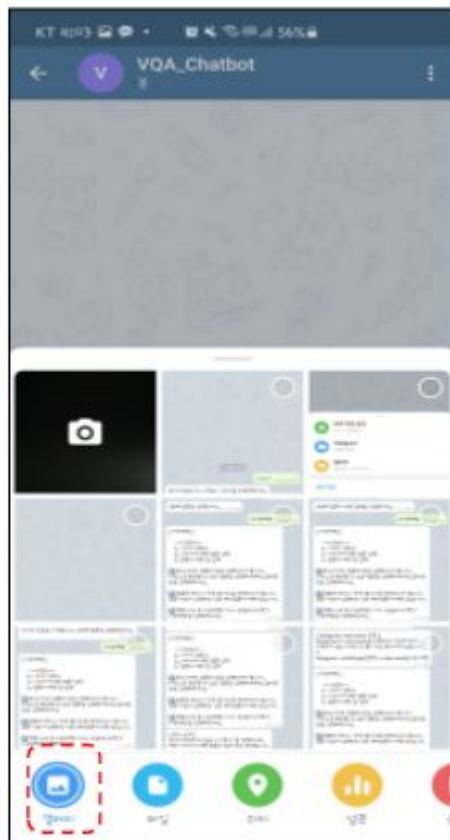
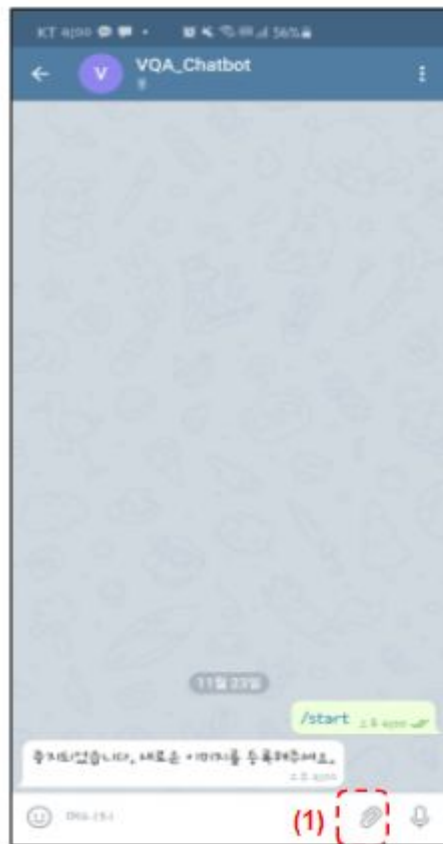
---

사용방법

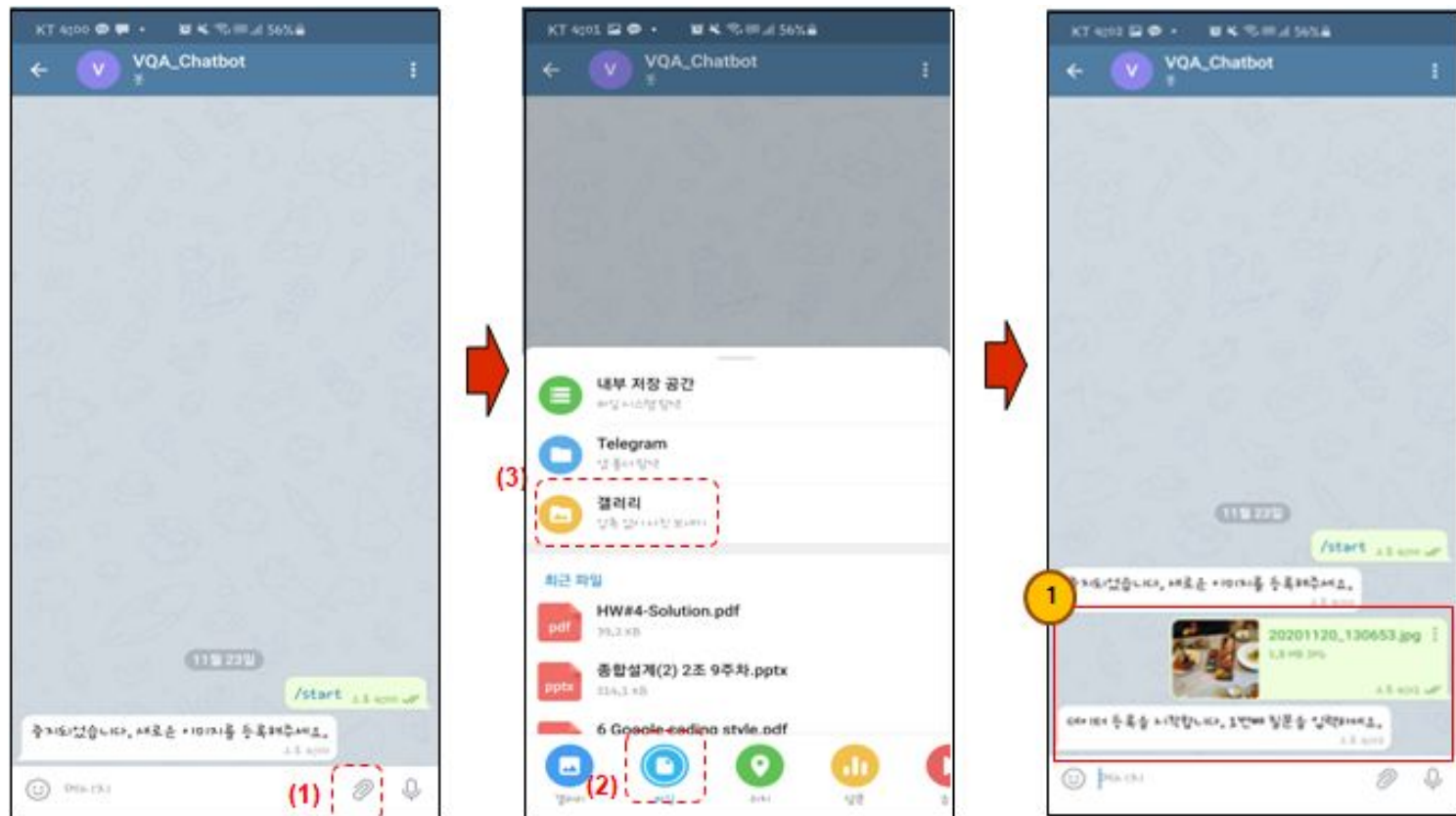




















  
\_AgAD\_QEAArFE  
uVU\_Ryu\_22.jpg


  
\_AgADBgIAAiors  
FU\_KimYounghe  
e.jpg


  
\_AgADiQIAAnz0  
2VU\_hongik\_test  
2.jpg


  
\_AgADiwEAAgzj  
qFU\_hongik\_test.  
jpg


  
\_AgADiwIAAnz0  
2VU\_hongik\_test  
2.jpg


  
\_AgADjQEAAgzj  
qFU\_hongik\_test.  
jpg


  
\_AgADTAEAAqva  
uFU\_hongik\_test.  
jpg


  
\_AgADwQEAAAn  
OoUVU\_hongik\_  
test.jpg


  
\_AgADxQEAAtvu  
2FU\_Hongik01.jp  
g


  
\_AgADyQEAAme  
mwVU\_Hongik0  
1.jpg


  
\_AQADDcn9a3Q  
AA8LVBgAB\_hon  
gik\_test4.jpg

  
\_AQADJb-ObHQ  
AA3e3BAAB\_Ryu  
\_22.jpg









파일명 : '\_unique\_id' '\_username'.jpg

5MB

결과 1개 나옴 (총 43건)

액션:   1 중 아무것도 선택되지 않았습니다.

<input type="checkbox"/>	IMAGE ID	IMAGE	WORKER ID
<input type="checkbox"/>	100	data/_AgADjQEAAgzjqFU_hongik_test.jpg	DATA_WORKER object (hongik_test)

04

---

결 론

## 문제점 및 해결방안

문제점	해결방안
사용자가 쓰레기 데이터를 무작위로 입력 가능	블랙리스트 or 인가된 사용자만 접근 가능 (現 텔레그램 사용자라면 누구나 등록 가능)
메타데이터가 포함된(압축되지 않은 파일) 이미지 전송할 때의 복잡한 절차	현재 주 목적은 사진, 질문, 답 데이터가 중요하므로 메타데이터가 포함된 사진만을 받는다면 사진 올리는과정이 복잡할 수 있음 → 메타데이터(위도, 경도 등) 정보가 없는 이미지 데이터를 받을 수 있음
사용자가 등록한 전체 이미지, 질문, 답변 수 확인	필요시 키워드 입력 후 등록된 개수 카운트 전송

## 기대효과

간편한 데이터 등록

장소에 제한없이 핸드폰으로 데이터 등록 가능

정확한 데이터 수집

데이터 등록에 대한 피드백을 통해  
정확한 데이터 수집이 가능

메타데이터 이미지

메타데이터(위도, 경도)가 포함된 이미지 데이터  
활용한 추가 프로젝트 구상

연구 기여

수집된 이미지의  
메타데이터를 활용한 추가연구

**THANK  
YOU**