

HUFS Graduation Project

OCR을 활용한 약력 관리 및 조회 서비스

작성일자 : 2021-03-08

201501730 송진호

201501114 류준형

Index

01 서론

02 관련 기술 조사 및 분석

서론

연구 배경 & 연구 목표



01

연구 배경

02

연구 목표



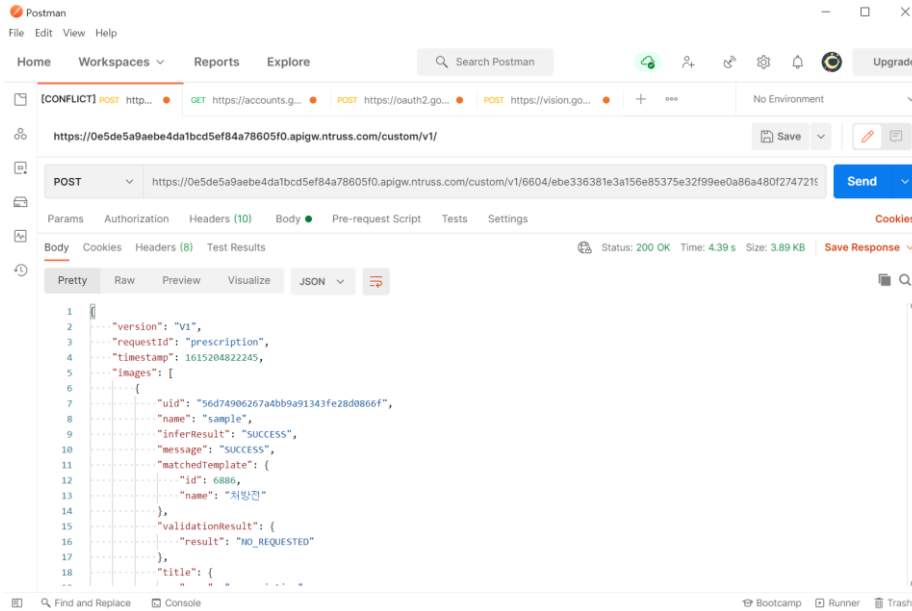
관련기술 조사 및 분석

OCR & 관련 서비스 및 API

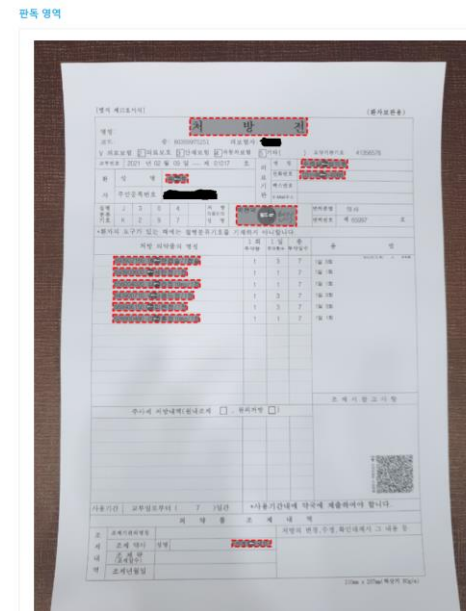
OCR

- ✓ OCR이란 Optical Character Recognition의 약자로 광학 문자 인식을 뜻한다.
- ✓ 사진으로 찍은 문서 또는 종이에 인쇄되었거나 수기로 쓴 문자, 기호 등에 빛을 비춰서 반사된 광선을 전기 신호로 바꾼 후에 컴퓨터에 입력하는 기술을 말한다.
- ✓ 보통 세금이나 공공요금 고지서 및 영수증에 많이 쓰이고 있으며, RPA(Robot Process Automation)와 같이 단순하고 반복적인 업무를 자동으로 처리해주는 소프트웨어에도 자주 사용된다.
- ✓ 시중에 서비스되고 있는 대중적인 OCR로는 'Naver CLOVA OCR', 'Google Cloud Vision API' 등이 있으며 무료로 제공되는 오픈소스 라이브러리로는 'Tesseract'가 가장 많이 알려져 있다.

Naver CLOVA OCR



[Postman 결과]



[Beta Test 결과]

처방전 자동분류 (처방전)

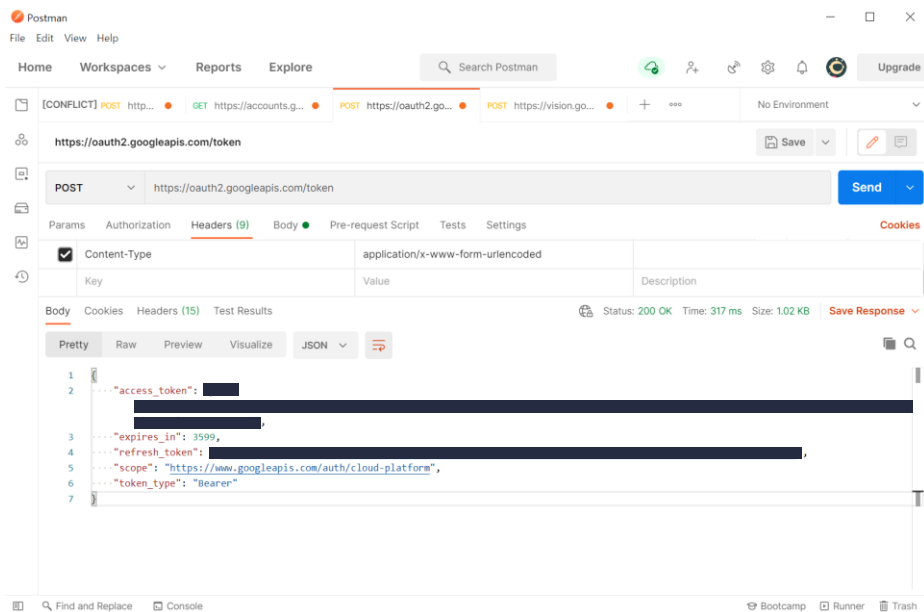
이름	내용
prescription	처방전

필드

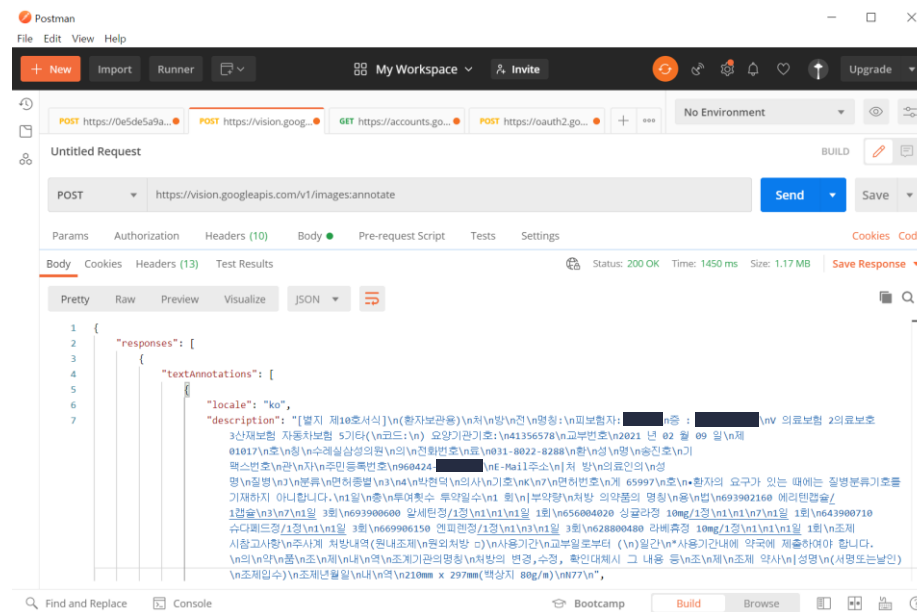
필드명	이름	내용
필드 01	user_name	송진호
필드 02	hospital_name	수레상사병원
필드 03	hospital_call	031-8022-8288
필드 04	doctor_name	
필드 05	pharmacy_name	(서용되는날)
필드 06	pharmacist_name	
필드 07	compound_date	박현덕 박현덕
필드 08	medicine_name1	693902160 메리텐캡슐/1캡슐
필드 09	medicine_name2	693900600 알세린정/1정
필드 10	medicine_name3	656004020 상글라정 10mg/1정
필드 11	medicine_name4	643900710 슈다펙도정/1정
필드 12	medicine_name5	669906150 엔피렌정/1정
필드 13	medicine_name6	628800480 라베후정 10mg/1정
필드 14	medicine_name7	
필드 15	medicine_name8	
필드 16	medicine_name9	
필드 17	medicine_name10	
필드 18	medicine_name11	
필드 19	medicine_name12	
필드 20	medicine_name13	

- ✓ Google과는 달리 설명을 읽는데 이질감 없이 사용법을 쉽고 자세하게 설명했다.
- ✓ 이미지 전체에 대해서 OCR을 적용하는 방법과 특정 구간별로 OCR을 적용하는 방법 2가지를 모두 사용 가능하다.
- ✓ 실험해본 결과, 한글과 영어를 가리지 않고 모두 100%의 인식률을 보여줬으며 필기체 또한 매우 잘 인식하였다.
- ✓ 회원가입 후 무료로 이용할 수 있는 기간이 있으며, 인식 구간을 많이 설정할수록 청구 금액이 증가하게 된다.

Google Vision API



[OAuth 2.0 인증]



[Google OCR 적용]

- ✓ 기본적으로 Google의 API를 사용하려면 OAuth 2.0 인증을 받아야 해서 API를 사용하기 위한 절차가 복잡하다.
- ✓ 또한, OAuth 2.0 인증 절차가 모두 영어로 되어있고 API 사용법 또한 번역체로 되어있어서 사용자 편의성이 낮다.
- ✓ 이미지 전체에 대해서 OCR을 적용하는 방식만 제공하고 있어서 필요한 부분만 추출하는 것이 불가능하며, 추가적인 코딩 작업 필요할 것으로 생각된다.

기타 OCR

- 사이냅소프트(SYNAP OCR)

- ✓ 인식 데이터를 통해서 지속적으로 학습이 가능한 딥러닝 기반의 OCR
- ✓ 자체 회전 보정 알고리즘을 사용하여 정확도 향상 가능
- ✓ 클라우드 기반의 서비스가 아닌 설치형 AI OCR 솔루션
- ✓ 인식된 글자에 마우스를 위치시켜서 OCR 적용 결과를 보여줌
- ✓ 학습되지 않은 글꼴 및 왜곡, 노이즈 등의 인식을 방해하는 요소에 강함
- ✓ 도메인 딕셔너리를 활용한 인식 정확도를 보정할 수 있음

- Free Online OCR

- ✓ 한글 및 필기체를 인식하지 못하는 단점이 존재
- ✓ OCR 결과를 Adobe PDF, Microsoft Word, Microsoft Excel, TXT, RTF, HTML 파일 등의 원하는 형식으로 변환 가능
- ✓ SOAP 및 REST 인터페이스를 사용하여 구현되며 HTTP 또는 HTTPS 요청을 통해서 API를 통해 액세스 가능

기타 OCR

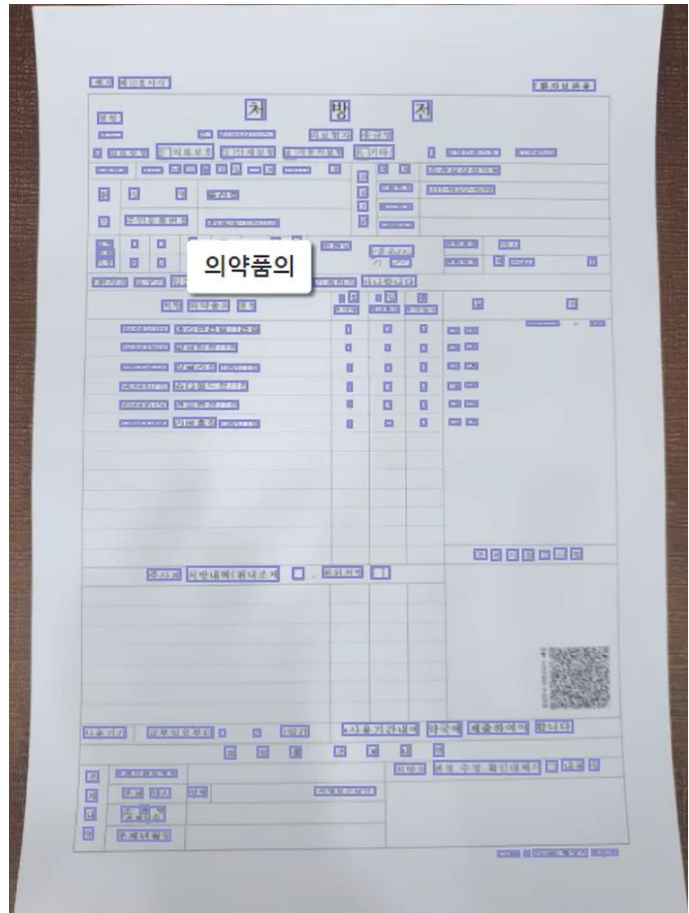
- ReTIA(ROSE OCR)

- ✓ 스캔한 문서 파일, 그림, 이미지 및 PDF 파일을 OCR 엔진을 적용하여 결과물을 워드 또는 엑셀 형식으로 변환해주는 무료 온라인 OCR 서비스
- ✓ 한글 인식률이 'Naver'와 'Google'에 비하여 좋지 않은 편이며, 필기체를 인식하지 못함
- ✓ 한글 외에 영어, 한자, 숫자를 추가로 인식 가능

- Easy OCR

- ✓ 네이버의 CRAFT를 Detection 모델로 사용하고 있고 CRNN을 기반으로 문자를 인식함
- ✓ 한글을 포함한 40여개의 언어를 지원하며 free-training 모델 또한 같이 제공하고 있음
- ✓ GPU 연산을 지원하기 때문에 처리속도가 빠름

기타 OCR



[SYNAP OCR]

11별지 제10호서식 (전자보관용)

처방전

코드: 4136578 중: 2021년 02월 09일 - 제 031-백 22-82

의약품명: 주혈장정액의 a

환자명: 송진호

의사명: 박현덕

의사번호: 65997

환자의 요구가 있는 때에는 질병분류기호를 기재하지 않습니다.

처방 의약품의 명칭	량	단위	용법
693902160 에리덴 1g 용/1 정	1	3	1일 331
69390 00 알레탄정/1 정	1	1	181 1st
656004020 실라정 10mg/1	1	1	181 1st
정	1	3	181 3st
643900710 슈다페드정/1 정	1	3	181 3st
669906150 엔피한정시정	1	1	1일 1st
628800480 라베유정 10mg/1 정	1	1	1일 1st

주사제

사용기간: 교부일로부터 (7) 사용기간내에 약국에 제출하여야 합니다.

의약품 2제

제명 (서명 또는 날인)

처방의 변경, 수정, 확인 대체시 그 내용

210x 297 (복합지 80g/2)

[ROSE OCR]

1 별지 제10호서식 (전자보관용)

처방전

의약품명: 주혈장정액의 a

환자명: 송진호

의사명: 박현덕

의사번호: 65997

환자의 요구가 있는 때에는 질병분류기호를 기재하지 않습니다.

처방 의약품의 명칭	량	단위	용법
633652165 에리 덴탈정/1 정	1	3	1일 331
633900600 알레탄정/1 정	1	1	181 1st
D82C 실라정 10mg/1	1	1	181 1st
43300/10 슈다페드정/1 정	1	3	181 3st
'6996 150 엔 피한정시정	1	1	1일 1st
628800480 라베유정 10mg/1 정	1	1	1일 1st

주사제

사용기간: 교부일로부터 1 r 7 1 E

의약품 2제

제명 (서명 또는 날인)

처방의 변경, 수정, 확인 대체시 그 내용

210x 297 (복합지 80g/2)

[Free Online OCR]