

OCR을 활용한 약력 관리 및 조회 서비스

1. 서론

1.1 연구 배경

현대 사회를 살아가는 사람들은 다양한 질병에 노출되어 있다. 건강에 이상이 생기면 병원에서 진료를 받은 후에 약국에서 약을 처방받고 일정 기간에 걸쳐서 복용한다. 의약 지식이 없는 대부분의 일반 사람들은 본인이 어떤 약을 처방받았는지에 대해서 크게 관심을 가지지 않고 병원에서 의사가 처방해준 그대로 복용한다. 약사는 ‘약사법 제24조(의무 및 준수 사항) 4항[1]’에 따라서 환자에게 처방 내용을 설명할 의무가 있다. 따라서 약사로부터 처방 내용에 관한 설명을 듣고, 처방전을 통해서 약에 대한 정보를 습득할 수 있어서 한동안은 무슨 약을 어떤 이유로 처방받았는지에 대해서 조금이나마 기억에 남을 수는 있다. 하지만 시간이 지나고 증상이 호전되어 복용을 중단하게 되면 그 당시에 무슨 이유로 정확히 어떤 약을 먹었는지에 대해서 기억하기란 쉽지 않다. 병원에서 진료를 받고 나면 ‘약국 제출용 처방전’ 외에도 ‘환자 보관용 처방전’을 따로 받을 수 있는데, 대부분은 이러한 사실을 모르고 있거나 받을 수 있음에도 관리하기 번거롭다고 느껴서 요구하지 않는다. 만약에 약사가 약을 잘못 지어 투약 사고가 발생했을 때를 대비하여 증거데이터로 사용하기 위해서는 처방전을 폐기하지 않고 일일이 모아야 하는데, 일어나지 않을지도 모를 만약의 사태를 대비하기 위해서 그것들을 모두 수집하는 것은 번거로운 일이 아닐 수가 없다.

건강보험심사평가원에서 제공하는 ‘내가 먹는 약! 한눈에[2]’ 서비스를 통해서 조회일 기준으로 최근 1년간 병원(약국)에 방문하여 조제 받은 의약품 투약 내역 및 개인별 의약품 알레르기 및 부작용 정보 등을 확인할 수 있다. 하지만 개인이 데이터를 조회하기 위한 과정에서 본인인증 및 공인인증서를 입력하는 등의 과정이 번거롭고 조회일 기준으로 최근 1년간의 데이터만 제공하기 때문에 조회할 수 있는 데이터가 제한적이라는 단점이 있다[3].

이러한 점에 기인하여 열람 데이터에 제한을 두지 않으면서 번거로운 과정을 거치지 않고 약력 데이터를 쉽게 조회할 수 있는 서비스를 기획하게 되었고, 기존에 건강보험심사평가원에서 제공하던 ‘내가 먹는 약! 한눈에’ 서비스를 보다 발전시켜서 약력 데이터를 편리하게 조회 및 관리하고자 하는 의도로 본 연구를 고안하게 되었다.

1.2 연구 목표

본 논문에서는 현재까지 복용 내역 등의 약력 데이터를 병원에 제출하거나 투약사고에 대비하기 위한 증거데이터를 수집하는 등의 개인적으로 필요한 상황이 발생할 때, 건강보험심사평가원을 통해서 복잡하고 번거로운 과정을 거치면서 데이터를 조회하지 않고, 사진 촬영한 처방전 내용을 OCR을 활용하여 인식 및 이를 데이터베이스화하여 어플을 통해서 쉽게 관리하고 조회하는 서비스를 개발하는 것을 목표로 한다. 연구를 진행함에 따라서 처방전 사진을 촬영하는 것만으로 개인의 건강상태를 추적할 수 있는 헬스케어 데이터 및 투약사고에 대비하기 위한 증거데이터를 편리하게 수집할 수 있을 것으로 기대된다.

2. 관련 기술 조사 및 분석

2.1 OCR(Optical Character Recognition)

OCR이란 Optical Character Recognition의 줄임말로 광학 문자 인식을 뜻한다. 풀어서 말하면 빛을 이용해서 문자를 판독하는 기술로 사진으로 찍은 문서 또는 종이에 인쇄되었거나 손으로 쓴 문자, 기호 등에 빛을 비춰서 반사된 광선을 전기 신호로 바꾼 후에 컴퓨터에 입력하는 장치이다. OCR은 보통 세금이나 공공요금 고지서 및 영수증에 많이 쓰이고 있으며, RPA(Robotic Process Automation)와 같이 단순하고 반복적인 업무를 자동으로 처리해주는 소프트웨어에도 자주 사용되고 있다. 이 내용은 2.3에서 보다 자세하게 다루기로 한다.

시중에 서비스되고 있는 대중적인 OCR로는 'Naver CLOVA OCR', 'Google Cloud Vision API', 'Kakao Vision API' 등이 있으며, 무료로 제공되는 오픈소스 라이브러리로 'Tesseract'가 가장 많이 알려져 있다. 본 연구에서는 처방전 규격을 통해서 문자를 인식할 계획이므로, OCR을 구간별로 적용이 가능한 'Naver CLOVA OCR'을 사용하기로 하였다.

2.1.1 Naver CLOVA OCR

판독 영역

The image shows a Korean medical prescription form. Red boxes highlight several key areas: the patient's name and ID number at the top, the doctor's name and ID number, and a list of prescribed medicines with their dosages. The form is titled '처방전' (Prescription) and includes fields for patient information, doctor information, and a list of medicines.

템플릿 자동분류 (처방전)

대표 샘플명

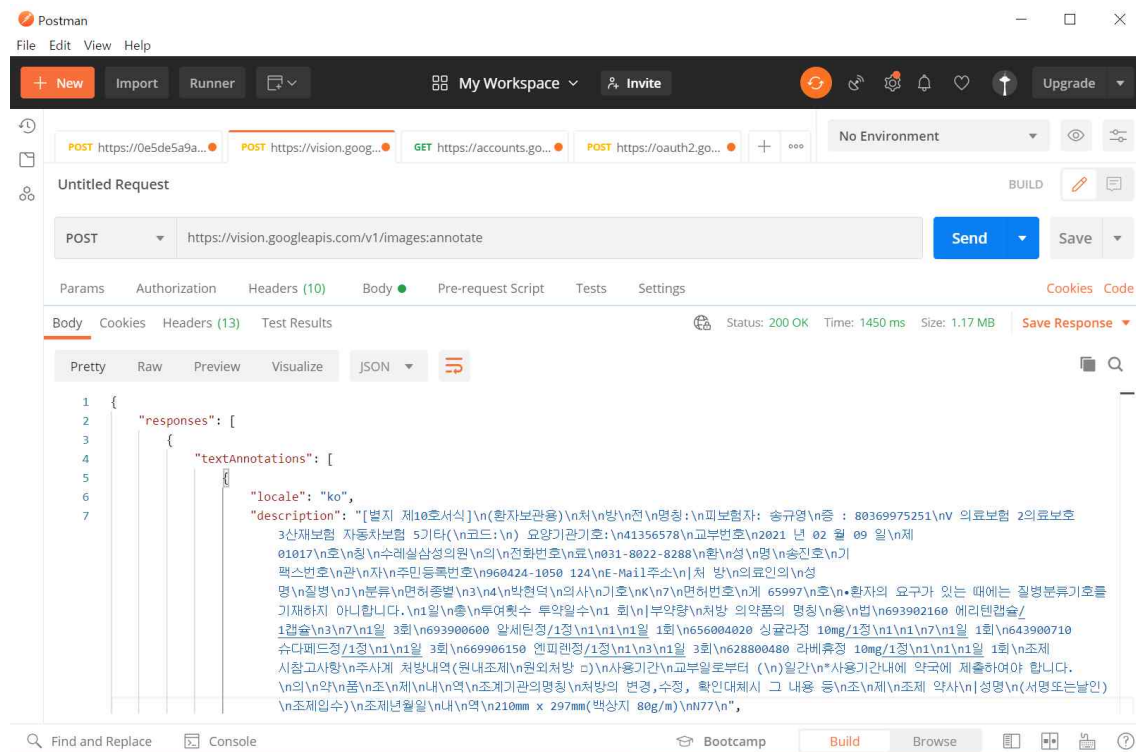
이름	내용
prescription	처방전

필드

판독 영역	이름	내용
필드 01	user_name	송진호
필드 02	hospital_name	수려심상성의원
필드 03	hospital_call	031-8022-8288
필드 04	doctor_name	
필드 05	pharmacy_name	(서명또는날인)
필드 06	pharmacist_name	
필드 07	compound_date	박현덕 박현덕
필드 08	medicine_name1	693902160 에리덴캡슐/1캡슐
필드 09	medicine_name2	693900600 알세틴정/1정
필드 10	medicine_name3	656004020 상글라정 10mg/1정
필드 11	medicine_name4	643900710 슈다팜드정/1정
필드 12	medicine_name5	669906150 엔피렌정/1정
필드 13	medicine_name6	628800480 라벡루정 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	

- 다른 외국기업과 달리 설명을 읽는데 이질감 없이 사용법을 쉽고 자세하게 설명했음
- 이미지 전체의 텍스트를 불러오는 다른 OCR과는 달리 구간별로 OCR을 지정할 수 있었음
- 회원가입 후 특정 기간만 무료 & 구간이 많아질수록 청구 금액이 증가하는 단점이 존재함
- 한글과 영어를 가리지 않고 100%의 인식률을 보여주며, 필기체 또한 매우 잘 인식함

2.1.2 Google Cloud Vision API



- 기본적으로 구글의 API를 쓰려면 OAuth 2.0을 통해서 인증을 받아야 해서 절차가 복잡함
- OAuth 2.0 인증 절차가 모두 영어로 되어있고, API 사용법도 번역체라서 해석이 어려웠음
- 이미지 전체의 텍스트를 불러오는 방식으로 필요한 부분만 추출하는 것에 있어서 번거로움

2.1.3 Kakao Vision API

- 애플리케이션 만들기 누르면 사용하기 편하게 API Key가 바로 나옴
- 구글, 네이버와 달리 직접 코드를 짜서 적용해보는 방법 외에는 OCR 테스트할 방법 X
- 구글, 네이버와 달리 'Object Storage'도 소개만 있고 어떻게 사용하는지에 대한 설명 X

2.1.4 사이냅소프트



- 딥러닝 기반
- 클라우드가 아닌 설치형 AI OCR 솔루션
- 인식된 글자에 마우스를 위치시키면 OCR 적용 결과를 보여줌
- 따로 응용할 수 있는 API는 없고 완성된 솔루션으로 보임(CLOVA처럼 전체 인식 결과를 출력하지는 못하는가?)

2.1.5 Free Online OCR

1 별지 제10호서식 1 (전자보화문)

4850

처 방 천

종 80889905251 과보철가 송규영

철 의로보호 미상재보철 D'용차보철 □'타 1 교양기판기 4 41369518 수레실상성

2021 년 02 월 00 일 제 01017

명 의철

환 C31-n

1006124

주변동복번호 ; 백 박현명 나軫g

제 17

환자의 요구가 있는 때에는 질병분류기호를

저방 외악종의 명 칭

633652165 에리 뎀랜슬/4슬

633900600 말세틴정//정 拓仄

D脉2C 심글라정 1點a/점

43300/10 슈다페 드정8/ 정

'69906 150 앤 11 랜 정 /1 정

628800480 라베 휴정 10mb/ 정

주사재 처방내역 (원내조제 □ 원외처방 □)

사용 149 ! 교부일로부터 1 r 7 1E

사용기간내에 약취제 * 하여야 합니다 .

조 세 내

제 21

21F6 887W (복상지 제한)

- 한글 및 필기체를 인식하지 못함
- 내용 추가
- 내용 추가
- 내용 추가

- 네이버의 CRAFT를 Detection 모델로 사용하고 있고, CRNN을 기반으로 문자를 인식함
- 내용 추가
- 내용 추가
- 내용 추가

2.2 ‘의약품 성분 약효 정보 조회 서비스’ Open API

2.3 관련 연구 및 기술을 사용한 사례 조사

3. 시스템 구성

1. OCR을 활용하여 처방전 내용을 데이터베이스화
2. 건강보험심사평가원에서 제공하는 ‘의약품 성분 약효 정보 조회 서비스’ Open API를 활용하여 의약품의 성분과 효능 및 주의사항 등의 기본적인 정보 제공[4]

4. 참조

- <https://www.law.go.kr/LSW/lsLinkProc.do?lsNm=%EC%95%BD%EC%82%AC%EB%B2%95&chrClsCd=010202&mode=20&ancYnChk=0> [1]
- <https://www.hira.or.kr/rg/dur/indvAgreeNew.do?pgmid=HIRAA050300000100> [2]
- <https://rankro.tistory.com/216> [3]
- <https://www.data.go.kr/data/15021027/openapi.do> [4]