

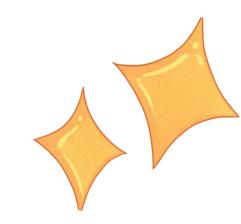


식품 알레르기 환자들을 위한 맞춤형 요리 어시스턴스 서비스 구현

팀명: ALARM(ALLERGY, ASSESSMENT, RECOGNITION, MANAGEMENT)

조장: 김상우

조원: 이은정, 이임형, 박인화



### 팀원소개



### DB,OCR



**상우** DB설계, 재료 전처리



**은정** 재 료전처리

### AI



**인화** RAG, 재료전처리



**B** RAG,재료 전처리

### 개발환경 및 라이브러리













**RAG** 







**CLOVA OCR** 

### 목차



- 1. 주제 및 배경 소개
- 2. 데이터 수집 및 전처리
- 3. DataBase
- 4. Chat 어시스턴스 서비스
- 5. 성분표 이미지인식 기능 구현
- 6. UI 구현
- 7. 서비스 시연
- 8. 향후 계획 및 느낀점

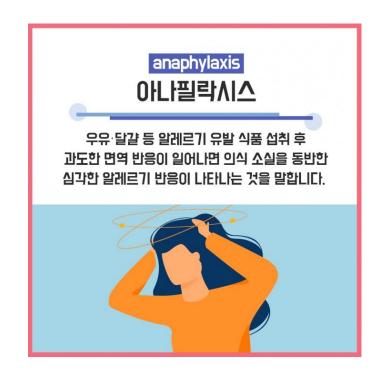


### 배경



#### 1. 식품 알레르기 위험성





(출처: HTTP://WWW.HEALTIP.CO.KR/NEWS/ARTICLEVIEW.HTML?IDXNO=3980#RS)

#### 3. 식품 알레르기 성분표 인식 저조

: 2015년 9월 전국 초, 중, 고 학생을 대상으로 한 조사에서 67.9%가 알레르기 유발식품 표시제에 대해 불만족. '성분표시라벨을 쉽게 찾을 수 없다', '성분표시에 대한 내용이 구체적이지 않다'.

#### 2. 식품 알레르기 표시대상 증가추세

- ㅇ「식품등의 표시기준」(식약처 고시)에 식품 알레르기 표시대상 22개 규정
- '03.5.23('04.5.23시행) : 표시 대상 알레르기 유발물질 총11개 규정
- \* 난류(가금류에 한함), 우유, 메밀, 땅콩, 대두, 밀, 고등어, 새우, 돼지고기, 복숭아, 토마토
- **'07.10.19**('09.4.30시행) : 새우 추가로 **총12개** 규정
- '11.11.7('13.1.1시행) : 아황산염 추가로 총13개 규정
- \* '12.12.31('12.12.31시행): 아황산염을 아황산류로 명칭 개정
- '15.4.8(15.48시행): 호두, 닭고기, 쇠고기, 오징어, 조개류(굴, 전복, 홍합 포함) 추가로 총21개 규정
- '18.4.26('20.1.1시행) : **잣** 추가하여 총 22개 규정

(출처: 식품의약품안전처)

#### 4. 벤치마킹(알러지알려줘, 알러드림)

- : 알러지알려줘 서비스 종료
- : 알러드림의 경우, HACCP(한국식품안전관리인증원)에 등록된 식품만 확인할 수 있음. 식품 표준 바코드만 허용하므로 접근성 하계

### 주제



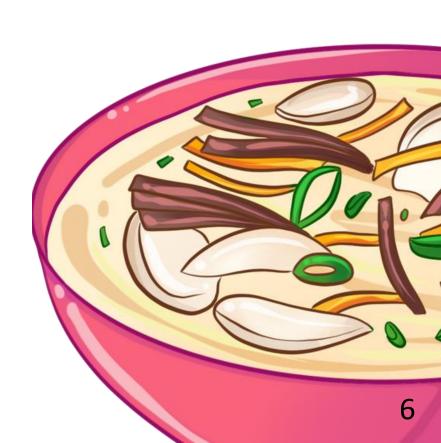
#### 식품 알레르기 환자들을 위한 맞춤형 요리 어시스턴스 서비스 구현

#### • 목표

o 식품 알레르기 환자들이 안전하고 건강한 식사를 할 수 있도록 돕는 통합 CHAT 어시스턴트 서비스를 제공

#### 기능

- o 맞춤형 요리 레시피 제공
- o 알레르기를 대체할 수 있는 재료 정보 제공
- o 제품성분표 라벨 확인을 위한 이미지 스캔 기능
  - 제품성분표의 작은 글씨 크기로 가독성 떨어짐
  - 식품 알레르기 제품 성분표 인식 저조 설문 결과
  - 따라서 가독성 높이는 방향으로 식품 알레르기 정보 제공하기 위함
- o 알레르기 증상 정보 제공
- o 교차반응 정보제공



# 전치리

# 데이터수집 및 전처리

### 데이터수집



레시피				
데이터명	출처			
조리식품의 레시피 DB	식품의약품안전처 - 공공데이터활용			
무료 레시피 데이터	KADX 농식품 빅데이터 거래소			

알레르기 기준				
데이터명	출처			
알레르기 유발식품 정의	국가법령정보센터(식품 등의 표시·광고에 대한 법률)			
알레르기 표시 정의	국가법령정보센터(소비자 안전을 위한 표시사항)			

대체식품			
데이터명	출처		
대체식품	부천구 어린이급식관리지원센터		

알레르기					
데이터명 출처					
알레르기 유발식품	국가법령정보센터(식품 등의 표시·광고에 대한 법률)				
	경기도 아토피·천식 교육정보센터				
교차반응 식품	allergyinsider				
	서울 서대문구 육아종합지원센터				

### 레시피 재료명 전처리



#### ✓ 문제

시간 소모 과다 - 2일간 진행 결과, 약 2000개 진행.

-> ChatGPT 를 응용하여 재료명 자동변환 <핵심 조건> 재료명 특징 추출, 입력-출력 1:1 보장

<모델> gpt-4turbo

```
prompt = f'''
Extract and normalize the following list of ingredients according to these conditions:

1. Choose one when ingredients are connected with '또는', '~나', 'or'

2. Use the same name for similar ingredients.

3. Ensure the core ingredient remains.

4. Make sure the number of ingredients in the output is the same as in the input list of ingredients

5. Ensure that the input and output ingredients are in a 1:1 correspondence

Provide a Dictionary in which original ingredient names are keys and normalized ingredient names are values.

Output must be korean

Just response Dictionary

{inputs}

...
```

### 레시피 재료명 전처리



### 전처리결과

	원재료명	정리후 재료명
20001	냉동새우(대)	새우
20002	올리고당약15g	올리고당
20003	양파 (중)	양파
20004	논우렁이	우렁이
20005	월계수잎생략 가능	월계수잎

### 시간 비용 비교

수작업	GPT 사용			
2일간 2000개	8시간 약 80000개			
125 / 1(시간)	10000 / 1(시간)			

# Chat 에시스턴스 서비스

### Chat 어시스턴스 서비스 구현 전체 흐름



입력: "나 ~~ 알레르기가 있어. ~~ 레시피 알려줘."

출력: 사용자 알레르기를 배제한 레시피 추천

#### 레시피 추천 기준

- 1) 원재료에 사용자 알레르기가 없어야 함
- 2) 가공식품에는 사용자 알레르기가 포함되어 있어도 레시피 추천 가능 단, 주의메시지 전달 (예: 가공식품 OO에 ㅁㅁ성분이 포함되어있습니다.)
- 1) 원재료와 가공식품 각각 사용자 알레르기와 교차반응 일으키는 성분에 대해 주의메시지 전달 (예: ☆☆식품의 △△성분은 사용자 알레르기와 교차반응을 일으킬 수 있습니다.)

### Chat 어시스턴스 서비스 구현 전체 흐름

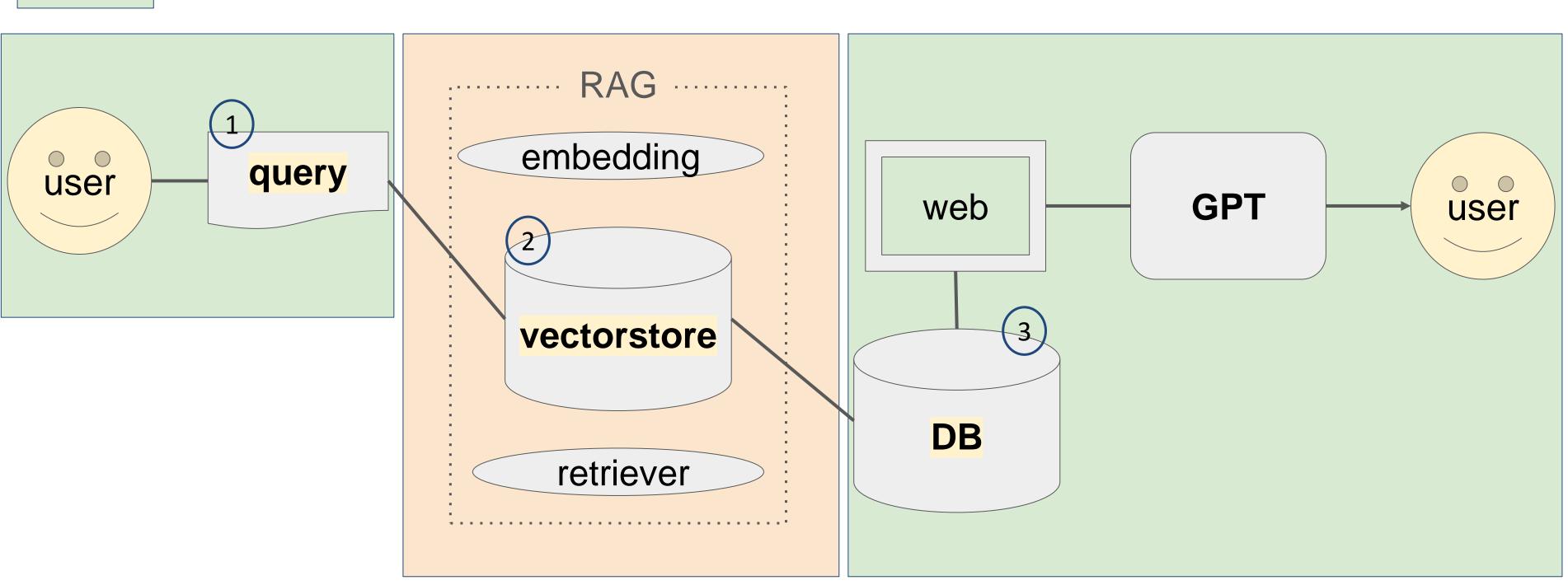


Part1

: 요리검색저장소

Part2

: 질의처리 및 응답생성



### Chat 어시스턴스 서비스 구현 전체 흐름



#### Part 1: 데이터 준비 및 벡터화

- [데이터 로드]
  - CSVLoader
- [텍스트 분할 및 벡터화]
  - RecursiveCharacterTextSplitter,
     HuggingFaceEmbeddings
- [벡터스토어 저장]
  - FAISS vectorstore

#### Part 2: 질의 처리 및 응답 생성

- [사용자 질의처리]
  - ChatOpenAI(gpt-4-turbo),
     retriever
- [알레르기\_교차반응매핑]
- [URL에서 레시피 정보 추출]
  - WebBaseLoader, BeautifulSoup
- [최종 레시피 및 경고 메시지 생성]
  - ChatOpenAI(gpt-3.5-turbo-0125)

Part 1. 데이터 준비 및 벡터화

### Part 1 데이터 준비 및 벡터화



Step 1: 요리 정보 로드

Step 2: 텍스트 데이터 분할

**CSVLoader** 

RecursiveCharacterTextSplitter

#### [출력물]

page\_content='레시피번호: 1

요리명: 어묵김말이

요리제목: 어묵김말이

url일련번호: 128671' metadata={'source': 'data/레시피이름\_url번호.csv', 'row': 0}

### Part 1 데이터 준비 및 벡터화



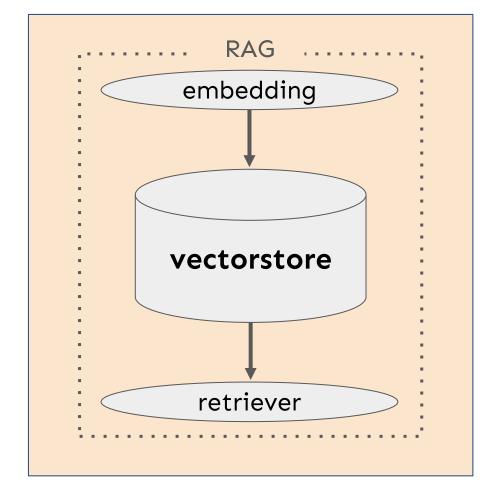
Step 3: 임베딩

HuggingFaceEmbeddings (BM-K/KoSimCSE-RoBERTa-multitask)

Step 4: 벡터스토어 저장

FAISS Index

[흐름도]

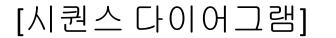


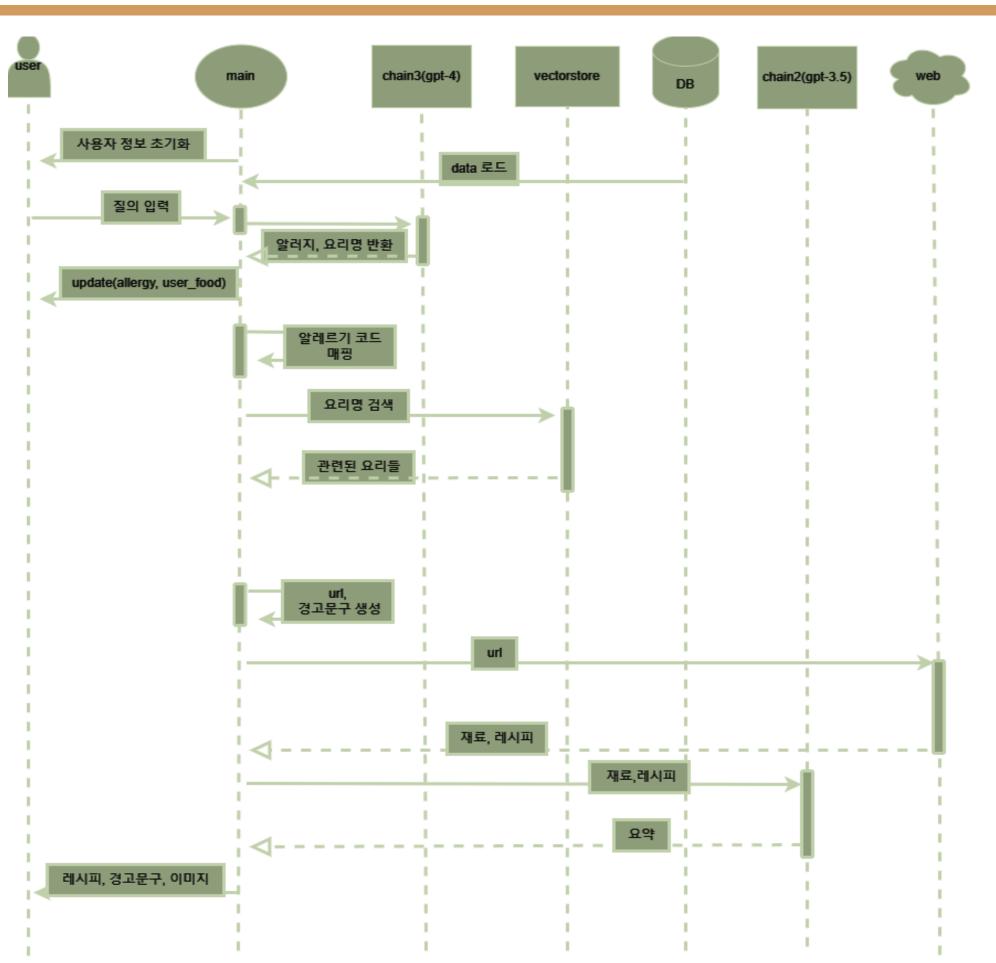
[출력물]

```
[[-0.01516053  0.00316845 -0.01526437 ... -0.00360666  0.00276193  0.00502264]
[-0.00879588 -0.00893074  0.0269921 ... -0.02946548 -0.02362109  0.03448407]
[-0.00919733 -0.03401578  0.02056984 ...  0.00403838 -0.00938999  0.00902816]
...
[ 0.01502976  0.01091959 -0.02008918 ... -0.02127218  0.01175904  0.01113749]
[-0.02181034 -0.01263297  0.03694781 ...  0.00602465  0.01433452  0.02345701]
[-0.01688495 -0.01953873  0.04405029 ...  0.00045326  0.01805663  0.01371677]]
```











#### [결과물]

요리제목: 그럴드 샌드위치

재료: 식빵, 가지, 토마토, 바질페스토(올리브오일), 모짜렐라치즈, 소금, 후추

#### 레시피:

- 1. 식빵 양면에 바질페스토를 발라준다.
- 2. 가지를 얇게 썰어 올리브오일과 소금, 후추에 버무려 팬에 굽는다.
- 3. 토마토도 썰어 팬에 굽는다.
- 4. 바질페스토를 바른 식빵 한 면에 구운 가지를 올리고 다른 한 면에 구운 토마토와 모짜렐라치즈를 얹는다.
- 5. 샌드위치 메이커에 넣고 굽는다. 샌드위치 메이커가 없으면 팬에 치즈가 녹을 때까지만 프라이팬에서 앞뒤로 굽는다.
- 6. 요플레와 계란 요리와 함께 곁들여 브런치 접시를 완성한다.

#### ※주의사항

사용자의 알레르기식품(['우유', '대두', '난류'])은 바질페스토(['우유']함유) 와 교차반응을 일으킬 수 있습니다!. 교차반응 모짜렐라 치즈(['우유']함유) 와 교차반응을 일으킬 수 있습니다!.

가공식품 식빵에 ['난류']가 함유 되었습니다!.

가공식품 바질페스토에 ['우유']가 함유 되었습니다!.

가공식품 모짜렐라 치즈에 ['우유']가 함유 되었습니다!.

가공식품





[Step 2 출력물]

[Document(metadata={'source': 'data/레시피이름\_url번호.csv', 'row': 108944}, page\_content='레시피번호: 110907 n요리명: 아침케이크\n요리제목: 아침케이크:밥신 빵을 짓다\nurl일련번호: 6933454'), Document(metadata={'source': 'data/레시피이름\_url번호.csv', 'row': 110766}, page\_content='레시피번호: 112735 n요리명 크림치즈토스트\n요리제목: 크림치즈토스트 : 통식빵 속 안까지 크림치즈를 머금은 오늘의 추천간식 ♥\nurl일련번호: 6935524'), Document(metadata={'source': 'data/레피이름\_url번호.csv', 'row': 143546}, page\_content='레시피번호: 145633\n요리명: 크로와상샌드위치\n요리제목: 냉동생지 크로와상 샌드위치 점심저녁도시락 빵간식 에프라이어요리\nurl일련번호: 6971725'), Document(metadata={'source': 'data/레시피이름\_url번호.csv', 'row': 46508}, page\_content='레시피번호: 47968\n요리명: 보로\n요리제목: 소보로 만들기!! 빵에다 솔솔 뿌려서 빵만들기\nurl일련번호: 6857465')]



Step 3: DB에서 재료코드 알레르기 여부 확인 및 주의 메시지 저장

DB

[원재료]

[가공식품]

[주의메시지]

```
1번째
레시피 번호 : 90930
원재료 확인 시작
I00668
I02418
I04283
I01499
I00608
T00006
I01301
레시피 번호 90930에 재료 101301에 알레르기 성분 존재
다음 번호 확인
2번째
레시피 번호 : 122165
원재료 확인 시작
```

```
레시피 번호 : 78801
                      'I00668': [],
원재료 확인 시작
                      'I00064': [],
I00668
                      'I01499': [],
100064
                      'I02709': ['A10'],
I01499
                      'I01976': ['A10'],
102709
                      'I00468': [],
I01976
                      'I00337': [],
I00468
I00337
가공식품 확인 시작
I00668
                      'I00668': ['A01'],
100064
                      'I00064': [],
I01499
                      'I01499': [],
102709
                      'I02709': ['A10'],
I01976
                      'I01976': ['A10'],
100468
                      'I00468': [],
100337
                      'I00337': [],
문제 없음
```

```
1) 교차반응(원재료, 가공식품)
```

:{'바질페스토': ['우유'], '모짜렐라 치즈': ['우유']}i

#### 가공식품(알레르기 성분 함유)

ː{'식빵': ['난류'], '바질페스토': ['우유'], '모짜렐라 치즈': ['우유']}:

원재료에서 사용자 알레르기 미포함 -> 레시피번호 저장 및 가공식품 확인



### Step 4: URL 생성 후 웹로더

WebBaseLoader

	URL_Num	Recipe_Num
76983	6896654	78801

https://www.10000recipe.com/recipe/6896654

[사이트]





### Step 5: 사용자에게 전달

[결과물]

GPT-3.5-turbo-0125

요리제목: 그렇드 샌드위치

|재료: 식빵, 가지, 토마토, 바질페스토(올리브오일), 모짜렐라치즈, 소금, 후추

#### 레시피:

- 1. 식빵 양면에 바질페스토를 발라준다.
- 2. 가지를 얇게 썰어 올리브오일과 소금, 후추에 버무려 팬에 굽는다.
- 3. 토마토도 썰어 팬에 굽는다.
- 4. 바질페스토를 바른 식빵 한 면에 구운 가지를 올리고 다른 한 면에 구운 토마토와 모짜렐라치즈를 얹는다.
- 5. 샌드위치 메이커에 넣고 굽는다. 샌드위치 메이커가 없으면 팬에 치즈가 녹을 때까지만 프라이팬에서 앞뒤로 굽는다.
- 6. 요플레와 계란 요리와 함께 곁들여 브런치 접시를 완성한다.

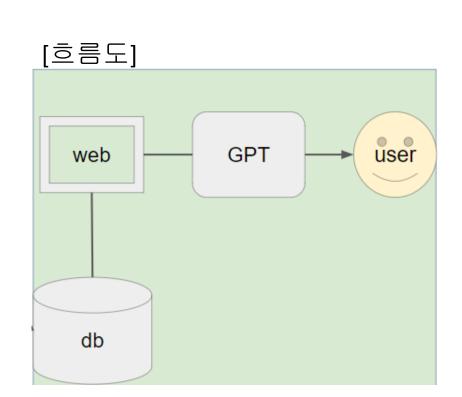
#### ※주의사항

사용자의 알레르기식품(['우유', '대두', '난류'])은 바질페스토(['우유']함유) 와 교차반응을 일으킬 수 있습니다!. 모짜렐라 치즈(['우유']함유) 와 교차반응을 일으킬 수 있습니다!.

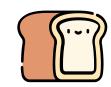
가공식품 식빵에 ['난류']가 함유 되었습니다!.

|가공식품 바질페스토에 ['우유']가 함유 되었습니다!.

|가공식품 모짜렐라 치즈에 ['우유']가 함유 되었습니다!.



### 예외처리1



#### <만약 요리가 존재하지 않는다면?>

- 1. 벡터스토어에 사용자의 요리명이 존재하지 않는 경우
  - : 요리명 자체가 없어서 검색 불가하므로
- 1. 선택된 모든 요리 원재료에 사용자 알레르기가 포함되는 경우
  - : 모든 요리를 검색했는데 더 이상 검색할 게 없으므로
- 1. 검색하는 문서의 개수(fetch\_k) 제한으로 못 가져오는 경우
  - : 요리가 있음에도 불가피하게 가져오지 못하므로

#### => gpt-3.5모델을 사용해서 해당 요리를 검색하도록 함

「죄송합니다. GPT 3.5를 사용해서 레시피 추천해드릴께요.

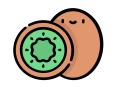
밀이 없는 르쁠라 레시피는 다양한 식재료로 대체할 수 있습니다. 예를 들어, 밀가루 대신 고구마가루나 옥수수가루를 사용하여 만들 수 있습니다. 아래는 밀이 없는 르쁠라 레 시피의 한 예시입니다.

#### 밀이 없는 르쁠라 레시피:

- 재료: 고구마가루 (또는 옥수수가루), 계란, 소금, 물
- 고구마가루를 그릇에 담고 소금을 조금 넣어 섞어줍니다.
- 계란을 풀어 고구마가루에 넣고 섞어줍니다.
- 물을 조금씩 넣어가며 반죽을 만들어줍니다.
- 팬에 식용유를 두르고 반죽을 부어 르쁠라 모양으로 만들어줍니다.
- 둘레가 노릇해질 때까지 뒤집어가며 익혀줍니다.
- 완성된 르쁠라에 원하는 장을 발라 맛있게 즐기세요.

밀 대신 다른 가루를 사용하여도 맛있는 르쁠라를 만들 수 있습니다. 즐거운 요리 시간 되시길 바랍니다!

### 예외처리2



#### 1. 질의에 알레르기가 없는 경우 (예: "카레 레시피 알려줘")

: 당연히 해당 요리명으로 검색 가능

#### 1. 질의에 "다른거 알려줘"라고 전달한경우

: 사용자가 이전에 입력한 알레르기와 요리 정보를 바탕으로 같은 요리의 다른 레시피 추천

: 없다면 랜덤하게 레시피 추천



# 성분표이미지인식기능구현

## 이미지 텍스트 추출

### OCR 모델선정



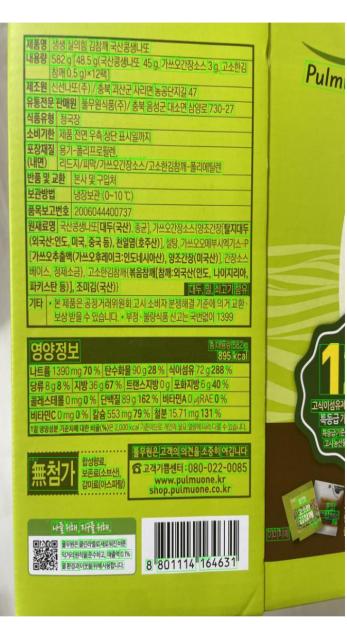
- 여러 OCR API중 한국어 인식이 가장 좋은 Naver CLOVA API로 사용
- Naver OCR 테스트결과, 화질이 나쁘지만 않다면 글씨를 깔끔하게 인식되어 읽혀짐

	Tesseract	EasyOCR	Google Vision	AWS Textract	Azure Document Intelligence	Naver Clova	Upstage	PaddleOCR
Open Source	0	0	X	Х	Х	Х	Х	0
한글 인식	중	하	중상	최하(지원X)	중상	상	상	중(추가 테스트 필요)
영문 인식	상	하	상	상	상	상	상	상
표+글자 인식	하	하 (유료 Enterprise용 별도 존재)	상	상(Only Eng)	상	중상	상	중
속도	중	최하	최상		중상	중하	중	하
특화 모델* 제공	х	x	0	0	O (영수증, 명함, 보험/세금 서류, 신분증 등)	O(영수증, 명함, 사업자등록증 등)	O(영수증)	х
Model Customization**	0	0	x	х	0	0	х	0
요금(1건)***	무료	무료	\$1.5 (매월 1000건까지 무료)		\$0.01 (종량제: 매월 500페이지 무료)	3원 (매월 300건 무료)	3원	무료
API 사용 난이도	하	하	중상	중	하	하	하	중
고객 지원	x	x	x	0	0	0	0	х
confidence score 제공	0	0	Δ	0	0	0	O (문서별,단어별 각각 제공)	0

<Original Image>



<Processed Image>



### 이미지 데이터 추출



#### • 이미지 글씨 추출 결과

제품명 생생 실의힘 김참깨 국산콩생나또 제조원 내용량 참깨 신선나또(주) 582 0.5 g [48.5 g)x12팩] / g(국산콩생나또 충북 괴산군 사리면 45 농공단지길 g, 가쓰 오간장소스 47 3 g, 고소한김 Pulm 유통전문 판매원 풀무원식품(주)/ 충북 음성군 대소면 삼양로 730-27 식품유형 청국장 소비기한 제품 전면 우측 상단 표시일까 지 포장재질 용기-폴리프로필렌, (내면) 리드지/피막/가쓰오간장소스/고소한김참 깨-폴리에틸렌 반품 및 교환 본사및 구입처 보관방법 냉장보관 (0~10°C) 품목보 고번호 2006044400737 원재료명 국산콩생나또[대두(국산), 종균], 가쓰오간장소 스{양조간장[탈지대두 (외국산:인도, 미국, 중국 등), 천일염(호주산)], 설탕, 가쓰 오오메부시엑기스-P[가쓰오추출액(가쓰오후레이크:인도네시아산), 양조간장(미 국산)], 간장소스 베이스, 정제소금}, 고소한김참깨{볶음참깨[참깨:외국산(인도, 나 이지리아, 파키스탄 등)], 조미김(국산)} 대두, 밀, 쇠고기 함유 기타 - 본제품은 공 정거래위원회 고시 소비자 분쟁해결 기준에 의거 교환 보상 받을 수 있습니다. 부정 - 불량식품 신고는 국번없이 1399 영양정보 총 내용량 582 g 895 kcal 1 나트 륨 1390 mg 70 % 탄수화물 90 g 28 % 식이섬유 72 g 288 % 당류 8 g 8 % 지방 36 g 67 % 트랜스지방 g 포화지방 6 g 40 % 콜레스테롤 mg 0 % 단백질 89 g 162 % 비타민A 0 gRAE 0 % 고식이섬유저 비타민C 0 mg O % 칼슘 553 mg 79 % 철 분 15.71 mg 131 % 특등급기 1일 영양성분 기준치에 대한 비율(%)은 2,000 kcal 기준이므로 개인의 필요 열량에 따라 다를 수 있습니다. 특등급 기준 고시 농산물 풀무원은 고객의 의견을 소중히 여깁니다 합성향료, ±첨가 고객기쁨센터: 080-022-0085 보존료(소브산), 감미료(아스파탐) www.pulmuone.kr shop.pulmuone.co.kr 풀무원 고소한 풀무원 김참깨 가쓰오 간장소스 나를 지구를 3g 위해. 위해. 05g 이미지예 풀무원은 클린라벨로 새로워진 바른 먹거리 원칙을 준수하고, 매출액 0.1% 를환경과 이웃을 위해 사용합니다. 8 801114 164631

### Open AI를 이용한 데이터 요약



- Open AI를 이용하여 장문의 OCR내용 요약
- 원재료명, 알레르기 유발물질, 제조시설의 식품품목만 뽑을수 있게 답변형식 지정
- GPT message를 Python 딕셔너리 형태로 변환시켜 Data로 사용

```
#답변형식 지정
ans = """{'원재료명':['''insert here'''], '함유성분':['''insert here'''], '같은 제조시설에서 제조한 식품':['''insert
here''']}"""
# ChatGPT 호출
response = client.chat.completions.create(
   model = "gpt-4-turbo",
   response_format = {"type": "json_object"},
   messages = [
   {"role": "system", "content": "당신은 영양성분표의 원재료 및 함유, 알레르기를 분석하여 JSON 형식으로 출력하는 데 도움이 되는 조
력자입니다."},
   {"role": "user", "content": f'''{full_text}을 분석해 주세요.
   {ans}의 형식으로 말해주세요.
   원재료에 있는 함유성분은 ['계란','소고기','돼지고기','닭고기','새우','게','오징어','고등어','조개류','우유','땅콩','호
두','잣','메밀',
   '대두','복숭아','토마토','밀','이황산류']만 입니다. 원재료에 있는 내용만 검색하고 그외의 것은 생략해주세요
    없을시엔 표시하지 말아주세요'''},
   temperature=0
```

(원재료명': ['국산콩생나또[대두(국산), 종균]',
'가쓰오간장소스{양조간장[탈지대두 (외국산:인도, 미국, 중국 등), 천일염(호주산)], 설탕, 가쓰오오메부시엑기스-P [가쓰오추출액(가쓰오후레이크:인도네시아산), 양조간장(미국산)], 간장소스 베이스, 정제소금}',
'고소한김참깨{볶음참깨[참깨:외국산(인도, 나이지리아, 파키스탄 등)], 조미김(국산)}'],
'함유성분': ['대두', '밀'],
'같은 제조시설에서 제조한 식품': []}