

## 레시피 정보 데이터 처리

2019-14240  
김연정

- 
1. 프로젝트 개요
  2. 활용 데이터
  3. 데이터로부터 도출할 결과물
  4. 전처리 계획 및 도출 데이터
  5. 프로젝트 의의
- 

### 1. 프로젝트 개요

이번 프로젝트에서는 레시피 데이터를 활용하여 두 가지 결과물을 도출하고자 한다. 첫 번째는 ‘메뉴 추천 서비스’이고, 두 번째는 ‘음식 수식어 분석 결과’이다. 사용할 레시피 데이터는 두 종류이다. 이 두 데이터는 상이한 성격을 가지고 있다. 첫 번째는 공공기관에 의해 깔끔하게 요약되고 정리된 데이터이고, 두 번째는 실제 사람들의 언어 사용이 반영된 데이터이다. ‘레시피 데이터’라는 주제로 함께 묶일 수 있는 이 두 데이터는 프로젝트 결과물을 도출하는 데에 유의한 정보를 제공할 수 있으리라 기대한다.

### 2. 활용 데이터

#### 1) ‘조리식품의 레시피 DB’ 레시피 데이터

##### 1) 정보 및 규모

2021년 8월 3일 최종 수정된 ‘조리식품의 레시피 DB’ 데이터는 메뉴명, 조리방법, 요리종류, 중량(1인분), 열량, 탄수화물, 단백질, 지방, 나트륨, 재료정보, 만드는 법 이미지, 만드는 법 텍스트 등을 필드로 가진 데이터이다. 총 1318개의 레시피가 저장되어 있으며, json 파일이다.

##### 2) 수집방법 및 출처

해당 데이터의 출처는 식품의약품안전처 데이터활용서비스이다. 다음의 URL에 접속하여 ([http://www.foodsafetykorea.go.kr/api/openApiInfo.do?menu\\_grp=MENU\\_GRP31&menu\\_no=661&show\\_cnt=10&start\\_idx=1&svc\\_no=COOKRCP01](http://www.foodsafetykorea.go.kr/api/openApiInfo.do?menu_grp=MENU_GRP31&menu_no=661&show_cnt=10&start_idx=1&svc_no=COOKRCP01)) Open API를 통해 데이터를 다운받았다.

### 3) 샘플 데이터

```
{
  "COOKRCP01": {
    "total_count": "1318",
    "row": [
      {
        "RCP_PARTS_DTL5": "우유100g, 매생이 20g, 찹쌀 15g, 단호박 50g\n밤 30g, 녹차소금 1g, 멸치육수 150g",
        "RCP_WAY2": "끓이기",
        "MANUAL_IMG20": "",
        "MANUAL20": "",
        "RCP_SEQ": "350",
        "INFO_NA": "64.3",
        "INFO_WGT": "",
        "INFO_PRO": "3.7",
        "MANUAL_IMG13": "",
        "MANUAL_IMG14": "",
        "MANUAL_IMG15": "",
        "MANUAL_IMG16": "",
        "MANUAL_IMG17": "",
        "MANUAL_IMG18": "",
        "MANUAL_IMG19": "",
        "INFO_FAT": "1.4",
        "HASH_TAG": "",
        "MANUAL_IMG02": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_02.png",
        "MANUAL_IMG03": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_03.png",
        "RCP_PAT2": "밥",
        "MANUAL_IMG04": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_04.png",
        "MANUAL_IMG05": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_05.png",
        "MANUAL_IMG01": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_01.png",
        "MANUAL01": "1. 찹쌀은 미리 불려준다.",
        "ATT_FILE_NO_MK": "http://www.foodsafetykorea.go.kr/uploadimg/cook/10_00350_1.png",
        "MANUAL_IMG06": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00350_06.png",
        "MANUAL_IMG07": "",
        "MANUAL_IMG08": "",
        "MANUAL_IMG09": "",
        "MANUAL08": "",
        "MANUAL09": "",
        "MANUAL06": "6. 죽이 어느 정도 완성되면 대차낸 호박과\n매생이를 넣어 반죽을 만들어준다.",
        "MANUAL07": "",
        "MANUAL04": "4. 밥은 굵게 다져주어 고명으로 사용한다.",
        "MANUAL05": "5. 우유와 육수를 섞은 뒤 찹쌀을 끓여준다.",
        "MANUAL02": "2. 매생이는 물에 담가 반가와 이물질을 빼준다.",
        "MANUAL03": "3. 단호박은 작은 사각형으로 잘라 멸치육수에\n데치준다.",
        "ATT_FILE_NO_MAIN": "http://www.foodsafetykorea.go.kr/uploadimg/cook/10_00350_2.png",
        "MANUAL11": "",
        "MANUAL12": "",
        "MANUAL10": "",
        "INFO_CAR": "23.7",
        "MANUAL19": "",
        "INFO_ENG": "122.3",
        "MANUAL17": "",
        "MANUAL18": "",
        "RCP_NM": "매생이찹쌀떡",
        "MANUAL15": "",
        "MANUAL16": "",
        "MANUAL13": "",
        "MANUAL14": "",
        "RCP_PARTS_DTL5": "통삼겹 250g, 닭육수 400g\n구이소스 : 저염된장 15g, 꿀 30g, 청주 15g, 참기름 15g, 다진마늘 10g, 다진대파 5g, 통깨 10g\n샐러드 : 수삼 1개(20g), 알배추 35g, 당근 30g\n샐러드소스 : 식초 30g, 다진마늘 15g, 참기름 15g, 맛간장 15g",
        "RCP_WAY2": "끓이기",
        "MANUAL_IMG20": "",
        "MANUAL20": "",
        "RCP_SEQ": "351",
        "INFO_NA": "120.6",
        "INFO_WGT": "",
        "INFO_PRO": "14.8",
        "MANUAL_IMG13": "",
        "MANUAL_IMG14": "",
        "MANUAL_IMG15": "",
        "MANUAL_IMG16": "",
        "MANUAL_IMG17": "",
        "MANUAL_IMG18": "",
        "MANUAL_IMG19": "",
        "INFO_FAT": "35.7",
        "HASH_TAG": "삼겹살",
        "MANUAL_IMG02": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_02.png",
        "MANUAL_IMG03": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_03.png",
        "RCP_PAT2": "반찬",
        "MANUAL_IMG04": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_04.png",
        "MANUAL_IMG05": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_05.png",
        "MANUAL_IMG01": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_01.png",
        "MANUAL01": "1. 구이소스 재료를 이용해 소스를 만든다.",
        "ATT_FILE_NO_MK": "http://www.foodsafetykorea.go.kr/uploadimg/cook/10_00351_1.png",
        "MANUAL_IMG06": "http://www.foodsafetykorea.go.kr/uploadimg/cook/20_00351_06.png",
        "MANUAL_IMG07": "",
        "MANUAL_IMG08": "",
        "MANUAL_IMG09": "",
        "MANUAL08": "",
        "MANUAL09": "",
        "MANUAL06": "6. 양념에 재워진 통삼겹을 구워주고 샐러드를\n담아 완성한다.",
        "MANUAL07": "",
        "MANUAL04": "4. 샐러드재료는 굵게 채 썰고 냉수에 담가\n냉각의 맛을 증하시켜준다.",
        "MANUAL05": "5. 샐러드소스를 만들어 샐러드를 버무려준다.",
        "MANUAL02": "2. 통삼겹은 닭육수에 삶아 냉수에 행궈 물기를\n빼준다.",
        "MANUAL03": "3. 통삼겹에 만들어진 소스를 발라 준 뒤 10분\n정도 숙성 시켜준다"
      }
    ]
  }
}
```

## 2) ‘만개의 레시피’ 레시피 데이터

### 1) 정보 및 규모

2007년부터 2020년까지 레시피 요리 플랫폼 ‘만개의 레시피’에서 수집된 레시피 정보를 담은 데이터이다. 2021년 9월 29일에 최종수정 된 버전을 활용한다. 레시피 일련번호, 레시피 제목, 요리명, 등록자 ID, 스크랩수, 요리소개, 재료 등 18개의 필드를 가지고 있다. 126123개의 레시피를 담고있으며, 54.2MB(홈페이지 설명기준 69.9MB)의 크기를 가지는 csv파일이다.

### 2) 수집방법 및 출처

농식품 빅데이터 거래소의 다음 URL에서 다운로드 받았다.

(<https://kadx.co.kr/product/detail/0c5ec800-4fc2-11eb-8b6e-e776ccea3964>)

### 3) 샘플 데이터

RCP_SNO	RCP_TTL	CKG_NM	RGR_ID	RGR_NM	INQ_CNT	RCMM_CNT	SRAP_CNT	CKG_MTH_AC	CKG_STA_AC	CKG_MTRL_A	CKG_KND	AC	CKG_IPDC	CKG_MTRL_C	CKG_INBUN_I	CKG_DODF_N	CKG_TIME_N	FIRST_REG_DT
128671	여목김말이	여목김말이	skfo0701	꽃날	9502	6	65	튀김	간식	가공식품류	디저트			맛있는 김말이(재료)	여목 2인분	초급	60분 이내	2.007E+13
128892	두부애 고리가 달렸어요!	skfo0701	꽃날		5506	3	26	부침	일상	해물류	밀반찬	고리가 너-무 (재료)	두부 1.3인분	초급	30분 이내	2.007E+13		
128932	입안에서 툭툭툭	skfo0701	꽃날		6783	8	35	굽기	일상	해물류	밥/죽/떡	간단하게 만들(재료)	밥 1+12인분	초급	30분 이내	2.007E+13		
131871	★현미호두죽 현미호두죽	cds1117	햇님&별님		2695	0	8	끓이기	일상	쌀	밥/죽/떡	현미호두죽 (재료)	현미 4인분	초급	30분 이내	2.007E+13		
139247	부들부들 보름복이갈비	skfo0701	꽃날		6793	3	94	굽기	술안주	견어돌류	메인반찬	오늘은 집에서(재료)	복어 2인분	초급	60분 이내	2.007E+13		
149207	토마토스파게 토마토스파게	skfo0701	꽃날		12609	2	36	볶음	일상	가공식품류	면/만두	한번 만들어 (재료)	파스타 1인분	초급	30분 이내	2.007E+13		
151148	표고버섯탕수 표고버섯탕수	skfo0701	꽃날		15271	2	194	튀김	손님접대	버섯류	메인반찬	탕수육 먹음직(재료)	표고 2인분	초급	30분 이내	2.007E+13		
153040	달콤한 마늘홍갈릭치킨	skfo0701	꽃날		8073	0	37	튀김	손님접대	닭고기	메인반찬	요즘 갈릭이 (재료)	닭 1마 2인분	중급	60분 이내	2.007E+13		
221094	바삭바삭 양파양파튀김튀김	skfo0701	박꽃나래		17346	14	239	튀김	간식	채소류	디저트	완전 바삭바삭(재료)	양파12인분	초급	30분 이내	2.007E+13		
221097	참치 삼각김밥 참치삼각김밥	skfo0701	박꽃나래		47046	8	348	기타	도시락	가공식품류	밥/죽/떡	맨날 편의점어(재료)	밥 2공 2인분	초급	30분 이내	2.007E+13		
221098	해물줄 가둬! (복배기해물떡)	skfo0701	박꽃나래		15531	8	34	볶음	술안주	해물류	밥/죽/떡	해물줄 가둬기(재료)	떡볶이 2인분	중급	30분 이내	2.007E+13		

## 3. 데이터로부터 도출할 결과물

### 1) 메뉴 추천 서비스

다음과 같은 물음을 거쳐 얻은 사용자의 응답과 레시피 데이터를 활용하여 메뉴를 추천하는 서비스를 제공하고자 한다.

Q1	식사 메뉴를 고르십니까, 후식을 고르십니까?
처리	답변에 ‘후식’이 포함되어 있으면 후식 class의 레시피 object들에서 검색을 시작하고, ‘식사’가 포함되어 있으면 식사 class의 레시피 object들에서 검색을 시작하고 Q4로 바로 이동한다. (후식 class와 식사 class는 아래 4.전처리 계획에서 서술)
Q2	직접 요리를 하실 예정입니까, 그렇지 않습니까?(요리 또는 외식 입력)

처리	답변에 '요리'가 포함된 경우 Q3-1로, 그렇지 않으면 Q3-2로 이동한다. 이후 결과를 제공할 때 레시피를 제공할지의 여부를 선택하기 위함이다.
Q3-1	사용하고자 하는 재료가 있습니까? (없을 시 공백)
Q3-2	드시고 싶은 특정한 식재료가 있습니까? (e.g. 닭고기, 면, 표고버섯 등, 없을 시 공백)
처리	KoNLPy의 Kkma noun분석기를 활용하여 식재료 리스트를 추출하고, 레시피 object의 재료 field에 해당 재료가 있는 object만을 필터링한다. 입력한 재료가 다수일 경우, 우선 재료들을 and로 처리하되, 모든 재료를 활용하는 메뉴가 없을 시 재료를 or로 처리한다. 사용자가 입력한 재료를 활용하는 레시피가 없을 경우, 해당 재료를 사용하는 메뉴는 지원하지 않는다는 응답을 제공하며 재입력 또는 서비스 종료를 선택할 수 있도록 한다.
Q4	특정 식재료에 알레르기가 있습니까? 있다면 입력해주시요. (없으면 공백)
처리	KoNLPy의 Kkma noun분석기를 활용하여 식재료 리스트를 추출하고, 레시피 object의 재료 field에 해당 재료가 없는 object만을 필터링한다.
Q5	채식주의자이십니까? (예 또는 아니요 입력)
처리	예/네 등의 긍정 답변이 입력될 시, 레시피 object의 재료 field와 레시피 field에 고기류가 포함되지 않은 object만을 필터링한다.

이 과정을 통해 필터링된 objects 중 랜덤으로 3개씩 사용자에게 보인다. 직접 요리를 한다고 응답한 경우 레시피까지 함께 제공한다. 이후, '다른 메뉴 추천을 원하십니까?'를 질문하고 응답에 따라 추가로 답변을 제공할지 결정한다.

## 2) 음식 수식어 분석

'만개의 레시피' 데이터의 두 번째 필드인 '요리 소개'에서, 음식을 수식하는 수식어들에 어떤 것들이 있는지 분석한다. 이를 통해 사람들이 음식을 표현할 때에 어떠한 수식어가 가장 많이 사용되는지 알아본다. 나아가, 수식어와 조회수 사이의 연관관계를 분석한다.

또한 레시피 작성 년도에 따라, 사람들이 자주 사용하는 수식어에 변화가 있는지 알아본다.

## 4. 전처리 계획 및 도출 데이터

### 1) '조리식품의 레시피 DB' 레시피 데이터

- json 포맷이므로 json 라이브러리를 활용한다.

- 이미지는 처리하지 않으므로, 만드는 법 이미지를 나타내는 속성은 모두 삭제한다.

- 일련번호, 영양정보, 중량, 해쉬태그 필드는 사용하지 않으므로 이 필드들을 모두 삭제한다.

- 후식 class와 식사 class를 만들고, 요리 종류가 '후식'인 데이터는 후식 objects로 생성해두고 그 외의 데이터는 식사 object로 생성해둔다. 이 둘의 object를 레시피 object라 통칭할 것이다.

이 과정을 통해 식사/후식 여부로 구분된, 음식명, 재료, 레시피를 담고있는 object들을 만든다. 이

object들은 메뉴 추천 서비스에 활용될 것이다.

## 2) '만개의 레시피' 레시피 데이터

- csv 데이터이므로, csv라이브러리를 활용한다.
- 결과물 도출에 활용될 '레시피 제목', '조회수'와 '최초등록일시'를 제외한 필드는 모두 삭제한다.
- 결과물 도출을 위해서는 '최초등록일시'에서 연도 정보만 필요하므로, 연도만 추출하여 정보의 양을 줄인다.  
'레시피 제목'이 가장 간결하고 핵심적으로 음식을 설명하는 데이터라고 생각하였기 때문에 해당 데이터를 활용하기로 결정하였다.
- KoNLPy의 Okt 분석기를 활용하여, 레시피 제목 내용의 pos를 분석한다. 분석된 내용 중 adjective에 해당하는 요소를 추출한다.
- adjective로 분석된 요소의 활용형이 다르면 빈도 분석 시에 문제가 생길 수 있으므로, kkma의 형태소 분석을 통해 어근을 추출한다.
- 추출된 어근과 조회수, 연도 정보를 결합하여 적합한 자료구조에 저장해둔다.

이 과정을 통해 음식을 수식하는 형용사의 어근들을 모을 수 있다. 이러한 어근을 빈도수로 정렬한다. 연도 구분없이 정렬해보고, 이후에는 연도에 따라 정렬해본다. 또한 각각 어근의 조회수(그 어근을 레시피 제목에 포함하는 글의 조회수)들을 합하여, 어근에 따른 조회수를 구한다.

### KoNLPy 사용 예시)

#### a. 이상적인 분석 결과

```
okt.pos('담백하고 부드러운 병어조림')

[('담백하고', 'Adjective'), ('부드러운', 'Adjective'), ('병어', 'Noun'), ('조림', 'Noun')]

print(kkma.morphs('담백하고'))
print(kkma.morphs('부드러운'))

['담백', '하', '고']
['부드럽', 'ㄴ']
```

이 경우, '담백'과 '부드럽'을 저장

```
okt.pos('색다른 노버터 흑미월병')

[('색다른', 'Adjective'),
 ('노', 'Noun'),
 ('버터', 'Noun'),
 ('흑', 'Adverb'),
 ('미', 'Adjective'),
 ('월병', 'Noun')]

print(kkma.morphs('색다른'))

['색다르', 'ㄴ']
```

이 경우, '색다르'를 저장

```

okt.pos('깔끔하고 개운한 올리면')

[('깔끔하고', 'Adjective'),
 ('개운한', 'Adjective'),
 ('올', 'Modifier'),
 ('라면', 'Noun')]

```

```

print(kkma.morphs('깔끔하고'))
print(kkma.morphs('개운한'))

['깔끔', '하', '고']
['개운', '하', 'ㄴ']

```

이 경우, '깔끔', '개운'을 저장

## b. 자주 발생하는 문제

```

okt.pos('매콤한 야채비빔국수')

```

```

[('매콤', 'Noun'), ('한', 'Josa'), ('야채', 'Noun'), ('비빔국수', 'Noun')]

```

```

okt.pos('달콤하고 바삭한 커피번')

[('달콤하고', 'Adjective'),
 ('바삭', 'Noun'),
 ('한', 'Josa'),
 ('커피', 'Noun'),
 ('번', 'Noun')]

```

```

print(kkma.morphs('달콤하고'))
print(kkma.morphs('바삭'))

['달콤', '하', '고']
['바삭']

```

‘매콤한’을 ‘매콤하다’로 보지 않고, 처음부터 ‘매콤+’한’로 분석하고, ‘바삭한’도 ‘바삭+’한’으로 분석하였다.

이 경우 adjective 추출에서 제외되므로 문제가 발생한다.

그러나 해당 경우는 여러 데이터에서 공통적인 패턴으로 발생하는 것 같이 보인다. Noun에 ‘한’(Josa)이 연이어 오는 경우, Noun을 수식어 어근으로 포함할 수도 있을 것이다.

## c. 해당 방식의 다른 한계점

1. 명사가 수식하는 경우는 detect 불가 e.g. 초간단 참치볶음밥, 영양만점 단호박 찹쌀도너츠
2. 두 형용사가 이어오는 경우 e.g. 고소 담백한 초간단 참치마요, 쫄깃향긋한 달래꼬막무침
3. 처음부터 제대로 구별하지 못하는 경우

e.g.

```

okt.pos('향긋하고 새콤한 달래무침')

[('향', 'Noun'),
 ('긋하고', 'Verb'),
 ('새', 'Modifier'),
 ('콤', 'Noun'),
 ('한', 'Josa'),
 ('달래', 'Noun'),
 ('무침', 'Verb')]

```

실제로 프로젝트를 진행하며 이러한 한계점을 어떻게 극복할 수 있을지 여러가지 시도를 해보고자 한다.

## 5. 프로젝트 의의

우선, ‘메뉴추천 서비스’는 식사 메뉴를 고민하는 상황에 있는 사람들에게 도움이 될 수 있을 것

이다. 해당 서비스는 여러가지 음식 종류를 담은 데이터를 활용하므로, 사용자는 고려하지 못한 새로운 선택지를 제공받을 수도 있다. 직접 요리를 하는 경우에는, 사용하고자 하는 재료를 반영하여 결과를 제공받을 수 있고 레시피까지 함께 제공되므로 여기서 오는 효용도 있으리라 기대한다.

음식 수식어 분석을 통해서는 사람들의 실제 언어 사용 양상을 분석할 수 있을 것이다. 사람들이 음식을 표현하기 위해 어떤 수식어를 많이 사용하는지, 사람들의 이목을 끄는 데에 유의한 영향을 미치는 수식어가 있는지 분석하는 것은 음식 마케팅 분야에도 활용될 수 있을 것이라 생각한다. 또한 해당 분석은 13년에 걸친 데이터를 활용하므로, 시간에 따라 변화하는 언어 사용에 대한 시사점도 얻을 수 있으리라 기대한다.