

레시피 정보 데이터 처리 : 메뉴 추천 서비스와 음식 수식어 분석을 위하여

자유전공학부 2019-14240 김연정

-
1. 프로그램 실행 방법
 2. 프로그램 구성 요소 및 원리
 3. 프로젝트 계획서와의 비교
 4. 결과 분석
 5. 의의 및 한계
-

본 프로젝트는 두 가지 레시피 데이터를 활용하여 '메뉴 추천 서비스'와 '음식 수식어 분석 결과'를 도출하였다.

1. 프로그램 실행 방법

본 프로그램은 Colab 환경에서 Python으로 작성되었다. 프로그램을 실행하는 방법은 프로젝트 ipynb 파일의 코드셀을 차례대로 실행하는 것이며, colab에서 작성된 만큼 colab에서 실행하시는 것을 권장드린다. 사용자가 별도로 처리를 해야할 부분은 없으며, 필요한 라이브러리는 코드셀 내에 import 되어있다.

다만 섹션 5. 음식 수식어 분석에서 matplotlib를 활용하여 그래프를 그릴 때 한글 깨짐 현상이 발생하여, 폰트를 다운받아 적용하였다는 점은 언급할 만하다. 사용자 환경에 폰트가 다운로드 되는 것을 선호하지 않는다면 해당 셀은 실행하지 않고 도출된 그래프를 참고하도록 권장하는 바이다. 또한 섹션 3 후반부에서 food_modifier를 실행하는 데에 10분이 넘는 오랜 시간이 소요된다는 점을 미리 밝힌다.

프로그램은 다섯 가지 섹션으로 나뉘어있고, 첫 번째 섹션에서는 데이터를 받아온다. 섹션 2, 3에서는 받아온 데이터에 전처리를 수행하고, 섹션 4, 5에서 그 데이터를 활용하여 결과물을 도출한다.

2. 프로그램 구성 요소 및 원리

프로젝트에서 활용한 데이터는 두 가지이다. 첫 번째는 '식품의약품안전처 데이터 활용 서비스'에서 제공하는 '조리식품의 레시피 DB' 데이터이다. 2021년 8월 3일 최종수정된 버전을 활용하였다. 이는 메뉴명, 조리방법, 요리종류, 중량(1인분), 열량, 탄수화물, 단백질, 지방, 나트륨, 재료정보, 만드는 법 이미지, 만드는 법 텍스트 등을 필드로 가지고 있다. 총 1318개의 레시피가 저장되어 있으며, json 파일이다.

두 번째는 레시피 요리 플랫폼 '만개의 레시피'에서 2007년부터 2020년까지 수집된 레시피 정보를 담은 데이터이다. '농식품 빅데이터 거래소'에서 제공되고 있으며, 2021년 9월 29일에 최종수정된 버전을 활용하였다. 이 데이터는 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.

데이터는 모두 각 출처의 홈페이지에서 다운받을 수 있다. 그러나 사용자에게 직접 파일을 다운받아 프로그램을 실행하라고 할 수는 없으므로, 본 프로젝트에서는 open API와 Github을 활용하여 프로그램이 직접 데이터를 다운받을 수 있도록 구성하였다.

데이터는 json, csv 형식을 갖고 있지만 둘 다 pandas library의 데이터프레임을 활용하여 처리하였다. 우선, 섹션2에서 진행한 '식품의약품안전처'의 '조리식품의 레시피 DB' 전처리를 간략히 설명하겠다.

2.1) 메뉴 추천 서비스용 데이터 전처리 (섹션 2)

json library를 통해 불러온 데이터는 필요한 정보만을 추출하여 data frame으로 변환하였다. 이는 재료, 메뉴 이름, 조리 방법 순서 등을 column으로 갖고 각 메뉴를 row로 가지고 있으며, 데이터 전 처리는 크게 '필요 없는 열 삭제'와 '각 row의 데이터 정제'로 이루어졌다.

필요없는 열 삭제 : 이미지, 일련번호, 해쉬태그, 영양정보 등은 분석에 활용되지 않으므로 이 정보를 가지는 열은 삭제되었다. 또한 열들의 이름과 순서를 알아보기 쉽게 변경하였다.

데이터 정제 : 재료 정보에 포함되어 있지만 식재료로 분류되기 어려운 정보들과 \n을 데이터에서 삭제하였다. 두부 110g, 쌀 90g 등과 같이 재료 정보를 나타내는 열에 '채소 준비', '초밥' 등과 같이 재료로 분류되기 어려운 내용들이 포함되어 있어, regular expression과 replace 함수를 활용하여 이들을 제거하였다. 또한, 조리 방법을 순서대로 담은 MANUAL01~20 열에서, 문장 내에 위치한 \n과 문장 끝에 붙어있는 의미 없는 알파벳을 삭제하였다. 이러한 정제 작업은 메뉴 추천 서비스에서 특히 결과물을 출력할 때 가독성을 높이는 이점을 가져왔다.

2.2) 음식 수식어 분석용 데이터 전처리 (섹션 3)

csv파일에서 pandas의 data frame으로 변환된 해당 데이터는 여러가지 열을 갖고 있었는데, 이중 레시피 제목, 조회수, 최초등록일시만을 남기고 다른 열은 삭제하였다. 이후 전처리 과정에서는 레시피 제목을 Mecab으로 분석하여 레시피 제목에 어떠한 음식 관련 수식어가 활용되었는지 알아내고자 하였다.

이를 위해 우선 Mecab 형태소 분석기가 활용하는 형태소 tag 정보들과, 데이터 중 일부를 샘플로 분석해본 결과를 통해 수식어가 가질 수 있는 tag의 후보를 도출하였다. tag의 후보로는 XR 어근, NNP 고유명사, NNG 일반명사, IC 감탄사, MM 관형사, VA 형용사, MAG 일반부사, 이렇게 7가지가 있었다. 이 tag들이 실제 음식 수식어로 기능할 수 있는지 검증해 본 후 활용하기 위해, 각각의 후보들은 10000개의 데이터로 검증 절차를 거쳤다. 10000개의 데이터에서 특정 tag를 가지는 형태소를 추출한 후, 이것이 음식을 수식하는 데에 활용되는지 직접 확인하였다. 이 과정은 ipynb 내 코드와 주석에 더욱 자세하게 기술되어 있다.

검증을 통해, 본 프로젝트에서는 1) XR 어근 2) ETM, EC, EF 앞의 VA 형용사 3) MAG 일반 부사, 이

세 가지를 수식어로 활용할 수 있을 것이라 결론 내렸다. 따라서 약 12만개에 달하는 전체 레시피의 제목을 형태소 분석한 후, XR, VA, MAG tag를 가지는 형태소를 수식어로서 추출하였다. 추출된 형태소는 추후 분석을 위하여, 그 형태소가 추출된 제목을 가진 글의 조회수, 작성 시기와 함께 새로운 data frame인 food_modifier에 저장되었다.

2.3) 메뉴 추천 서비스 결과 도출을 위해 구현한 작업 (섹션 4)

메뉴 추천 서비스는 사용자의 입력을 받아 그에 따른 처리를 해야하는 서비스이다. 따라서 우선 사용자의 응답이 긍정인지 부정인지 체크하는 함수를 만들었고, 사용자의 입력이 처리된 결과를 인자로 받아 메뉴를 추천하는 함수를 정의하였다.

메뉴를 추천하는 함수에서는 사용자의 입력에 따라 데이터를 필터링한다. 원하는 식재료 (ingredient)가 입력되면 이 재료를 ‘재료 정보’에 포함한 메뉴만을 뽑아온다. 알레르기 반응을 유발하는 식재료가 입력되거나, 고기 섭취를 원치 않는 채식주의자의 경우, 특정 식재료를 ‘재료 정보’에 포함하지 않는 메뉴만을 추가적으로 필터링한다. 필터링이 완료된 데이터가 도출되면, 사용자에게 이 데이터로부터 메뉴를 3가지씩 추천해준다.

사용자는 recommend_service 함수를 실행하여 메뉴 추천을 받을 수 있다. 함수는 사용자에게 다음과 같은 질문을 묻고 사용자로부터 적절한 답변을 얻어 처리한다.

Q1	식사 메뉴를 고르십니까, 후식을 고르십니까?
처리	답변에 ‘식사’가 포함되어 있으면 meal을 True로 설정하고 ‘후식’이 포함되어 있으면 meal을 False로 설정한다. 둘 다 포함되어있지 않으면 다시 질문한다. meal=False이면(후식 추천을 원하면) 바로 Q4로 넘어간다.
Q2	직접 요리를 하실 예정입니까? (네 또는 아니요 입력)
처리	응답의 긍정/부정 체크 함수를 활용하여, 대답이 긍정이면 cook을 True로 유지하고 대답이 부정이면 cook을 False로 설정한다. 이는 이후 결과를 제공할 때 레시피를 제공할지의 여부를 선택하기 위함이다. 응답이 긍정도 부정도 아니면 다시 질문한다. cook이 True이면 Q3-1을, False이면 Q3-2로 넘어간다.
Q3-1	사용하고자 하는 재료가 있습니까? 여러 재료를 입력할 땐 ,(쉼표)로 구분하여 주세요. (없을 시 공백)
Q3-2	드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(쉼표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백)
처리	공백이 입력되지 않았다면 “,”을 기준으로 string을 split하고 list에 저장해둔다.
Q4	특정 식재료에 알레르기가 있습니까? 있다면 입력해주십시오. 여러 재료를 입력할 땐 ,(쉼표)로 구분하여 주세요. (없을 시 공백)
처리	공백이 입력되지 않았다면 “,”을 기준으로 string을 split하고 list에 저장해둔다.
Q5	채식주의자이십니까? (네 또는 아니요 입력)
처리	응답의 긍정/부정 체크 함수를 활용하여, 대답이 긍정이면 vegetarian을 True로 설정하고 대답이 부정이면 True를 유지한다. 응답이 긍정도 부정도 아니면 다시 질문한다. 메뉴 필터링에 필요한 사용자 입력이 완료되었으므로, 메뉴 추천 함수를 호출하여 추천

	결과를 받아온다.
Q6	다른 메뉴 추천을 원하십니까? (네 또는 아니요 입력)
처리	응답의 긍정/부정 체크 함수를 활용하여, 대답이 긍정이면 메뉴 추천 함수를 재호출하고 대답이 부정이면 서비스를 끝낸다.

2.4) 음식 수식어 분석을 위해 구현한 작업 (섹션 5)

음식 수식어 분석은 섹션 3에서 도출한 `food_modifier dataframe`을 활용하였고, 세 가지 시사점을 도출해내기 위해 노력하였다.

우선, 사람들의 이목을 끌기 위하여 레시피 제목에 사용된 음식 수식어 중 무엇이 가장 많이 사용되었는지 분석하고자 하였다. 이를 위해, `food_modifier` 내에서 특정 수식어('modifier')가 몇 번 등장하였는지 그 수를 세어 보고 많이 사용된 순으로 정렬해보았다. 그러나 섹션3에서 진행한 형태소 분석이 완벽하지는 않았기에, 결과를 보며 음식 수식어가 아닌 것들은 직접 제거하였다.

다음으로, 수식어와 조회수에 관계가 있는지 알아보기 위하여 수식어 별로 (그 수식어가 사용된 제목을 가지는 글의) 조회수 평균을 구하였다.

마지막으로, 글이 작성된 연도에 따라 수식어 사용에 차이가 있는지 알아보기 위해 데이터를 'year'에 따라 그루핑하고, 연도별로 활용 빈도가 높은 수식어를 구하였다.

3. 프로젝트 계획서와의 비교

본 프로젝트는 계획서를 충실히 구현하기 위해 노력하였다. 대부분의 구현은 계획서와 동일하나, 실제 프로젝트는 크게 다음과 같은 차이점을 가진다.

먼저, 데이터의 형태소 분석에 `Kkma`가 아닌 `Mecab`을 활용하였다. 계획서의 피드백을 통하여 `Mecab` 분석기를 활용하는 것이 더 적합하고 좋음을 알게 되었고, 실제로 `Kkma`로 분석하는 것 보다 더 정확하고 자세하게 분석이 가능하였다.

다음으로, 메뉴 추천 서비스를 위한 데이터를 처리할 때 `class`를 만들어 처리할 것이라 계획하였는데 `pandas`의 `data frame`을 활용함으로써 `class`를 사용할 필요성을 느끼지 못하여 이를 사용하지 않았다. `json` 라이브러리와 `csv`라이브러리를 활용하겠다고 언급하였지만 이 또한 `pandas`를 대신 사용하게 되었다.

또한 계획서에서는, 메뉴 추천 서비스에서 사용자로부터 원하는 식재료나 제외할 식재료를 입력 받으면 이를 형태소 분석하여 리스트를 추출하고자 하였다. 그러나 이는 어려움이 있어 실제로 구현하지 못하였다. 만약 사용자가 제외할 식재료로 '닭가슴살'을 입력하고 이를 형태소 분석을 하면 '닭'과 '가슴살'이 되는데, 이 분석된 형태소를 기준으로 식재료를 제외하려면 닭가슴살보다

```
] mecab.pos('닭가슴살')
[('닭', 'NNG'), ('가슴살', 'NNG')]
```

훨씬 큰 범주인 '닭'까지 다 제외하게 되기 때문이다. 이같은 문제를 해결할 수 있는 좋은 방안이 아직 떠오르지 않았기에, 사용자 입력 그대로, 그리고 띄어쓰기가 제거된 입력을 형태소 분석 없이 데이터 필터링에 활용하였다.

마지막으로, 포함하고 싶은 식재료를 다수 입력하였을 때 이 식재료들을 우선 모두 포함하도록 `and`로 처리하고, 그 결과가 존재하지 않으면 `or`로 처리하고자 계획하였었지만, `or` 처리를 할 때 포함해야 할 식재료 간의 우선순위 혹은 기준을 정하는 데에 어려움을 겪어 `and`로 구현할 수밖에 없었다.

4. 결과 분석

4.1) 메뉴 추천 서비스는 사용자의 입력에 따라 다양한 메뉴와 레시피를 추천해주었다. 서비스에서 발생할 수 있는 다양한 결과는 아래에 예시로 제시되어있다. 계획했던 기능은 큰 문제없이 수행됨을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 1)

```
...
식사 메뉴를 고르십니까, 후식을 고르십니까? 음
다시 입력해 주세요. 식사 메뉴를 고르십니까, 후식을 고르십니까? 고민이네
다시 입력해 주세요. 식사 메뉴를 고르십니까, 후식을 고르십니까? 식사?
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 네!
사용하고자 하는 재료가 있습니까? 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백) 닭, 면
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시요. 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백) 없어유
채식주의자이십니까? (네 또는 아니오 입력) 아니요--
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:47: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

1. 월남쌈우동면 & 저염토마토소스
요리종류 : 반찬
재료 : 라이스페이퍼 10g(4장), 백김치 100g, 닭 가슴살 100g, 부추 50g, 우동면 200g, 마늘 10g(2개), 사과 80g, 양파 60g, 레드와인 10g(2작은술), 토마토 200g(1개), 핫 소스 10g(2작은술), 러
조리방법: 끓이기
레시피:
1. 마늘, 양파, 사과, 토마토는 다진다.
2. 팬에 기름을 두르고 마늘, 양파, 사과, 토마토 순으로 볶는다.
3. 2에 레드와인, 핫 소스, 레몬즙을 넣어 반으로 줄인 후 후추가루를 넣어소스를 완성한다
4. 우동 면은 삶은 후 체에 건져 물기를 뺀다.
5. 백김치, 부추는 3x1cm로 썬다.
6. 닭 가슴살은 삶아 찢는다.
7. 따뜻한 육수에 라이스페이퍼를 넣었다 꺼내어 우동 면, 백김치, 부추, 닭가슴살을 넣고 말아준다.
8. 토마토소스를 곁들인다.

2. 삼계부대찌개
요리종류 : 국&찌개
재료 : 재료 닭가슴살(100g), 당근(20g), 붉은 고추(6g), 풋고추(6g), 대파(6g), 썩갠(4g), 표고버섯(14g), 애호박(60g), 김치(40g), 팽이버섯(40g), 당면(21g), 숙주(40g), 떡(40g), 육수 당귀
조리방법: 끓이기
레시피:
1. 육수 재료를 넣고 20분간 끓여육수를 낸다.
2. 닭가슴살과 당근을 다져 모양을잡은 뒤 김 오른 찜기에 넣고 찐다.
3. 고추, 대파는 어슷 썰고, 썩갠, 표고버섯, 애호박, 김치는 먹기좋은 크기로 썰고, 팽이버섯은밑동을 잘라 손질한다.
4. 당면과 숙주는 뜨거운 물에 데쳐준비한다.
5. 양념장 재료를 섞은 뒤 육수(13g)를넣고 양념장을 만든다.
6. 준비한 재료를 넣고 육수를 부은 뒤양념장을 넣고 끓여 마무리한다.

실행 중(57초) Cell > recommend_service() > raw_input() > _input_request() > recv() > recv_multipart()
```

닭과 면을 모두 포함한 결과를 잘 도출해주고 있으며, 사용자의 다양한 응답도 프로그램에서 원하는 형식에서 크게 벗어나지 않는 한 어느정도 포용가능함을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 2)

```
...
식사 메뉴를 고르십니까, 후식을 고르십니까? 식사
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 네
사용하고자 하는 재료가 있습니까? 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백) 닭, 면
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시요. 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백) 토마토
채식주의자이십니까? (네 또는 아니오 입력) 아니요
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:47: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:63: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

1. 삼계부대찌개
요리종류 : 국&찌개
재료 : 재료 닭가슴살(100g), 당근(20g), 붉은 고추(6g), 풋고추(6g), 대파(6g), 썩갠(4g), 표고버섯(14g), 애호박(60g), 김치(40g), 팽이버섯(40g), 당면(21g), 숙주(40g), 떡(40g), 육수 당귀
조리방법: 끓이기
레시피:
1. 육수 재료를 넣고 20분간 끓여육수를 낸다.
2. 닭가슴살과 당근을 다져 모양을잡은 뒤 김 오른 찜기에 넣고 찐다.
3. 고추, 대파는 어슷 썰고, 썩갠, 표고버섯, 애호박, 김치는 먹기좋은 크기로 썰고, 팽이버섯은밑동을 잘라 손질한다.
4. 당면과 숙주는 뜨거운 물에 데쳐준비한다.
5. 양념장 재료를 섞은 뒤 육수(13g)를넣고 양념장을 만든다.
6. 준비한 재료를 넣고 육수를 부은 뒤양념장을 넣고 끓여 마무리한다.

2. 삼색 닭 만두국
요리종류 : 국&찌개
재료 : 닭가루 30g(2큰술), 당근 20g(6cm), 시금치 20g, 감자 20g(1/10개), 닭 가슴살 80g(2/3개), 대파 8g(10cm), 마늘 6g(2개), 청양고추 30g(3개), 숙주 40g, 당면 40g, 두부 50g(1/6모),
조리방법: 끓이기
레시피:
1. 육수 재료를 모두 넣고 물 600g과 소금 10g을 넣어 육수를 만든다.
2. 당근, 시금치는 다져 즙을 만들고, 감자는 삶아서 으갠다.
3. 닭가루에 당근 즙, 시금치 즙, 감자를 넣어 각각 반죽하여B 랩에 써서 30분 숙성시킨다.
4. 닭 가슴살, 대파, 마늘, 청양고추, 데친 숙주를 다진 후 물기를 없앤다.
5. 당면은 따뜻한 물에 불려 썰고 물기를 제거한다. 두부는 으개어 물기를 빼고, 양파와 양배추도 다진다.
6. 냄비에 물을 넣고 끓여 육수를 끓인다. 육수에 닭가슴살, 대파, 마늘, 청양고추, 숙주, 양파, 양배추, 당면, 두부를 넣고 끓인다.

실행 중(20초) Cell > recommend_service() > raw_input() > _input_request() > recv() > recv_multipart()
```

예시1과 동일한 조건이지만 알레르기 유발 식재료로 ‘토마토’를 추가하였다. 이에 따라, 예시1에서 추천 받은 메뉴인 월남쌈우동면 & 저염토마토소스는 잘 제거되었음을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 3)

```
식사 메뉴를 고르십니까, 후식을 고르십니까? 식사
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 아니오
드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(침표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백) 청경채
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시시오. 여러 재료를 입력할 땐 ,(침표)로 구분하여 주세요. (없을 시 공백) 닭
채식주의자이십니까? (네 또는 아니오 입력) 아니오
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:47: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:63: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

1. 표고버섯 청경채국
요리종류 : 국&찌개
재료 : 국물용 멸치 5g(3마리), 다시마 1장(5×1cm), 양파 10g(2×1cm), 표고버섯 기둥, 국간장 5g(1작은술), 물 300ml(1½컵), 청경채 20g(1개), 표고버섯 20g(2장), 다진 마늘 2g(1/3작은술)
조리방법: 끓이기

2. 비트를 곁들인 돼지 구이
요리종류 : 반찬
재료 : 삼겹 70g(5개), 건고추 5g(1개), 당근 30g(1/4개), 삼겹살 350g, 마늘 5g(1개), 설탕 30g(2큰술), 양송이버섯 70g(3개), 오렌지주스 400g(2컵), 밀계수윌 2g(1개), 청경채 70g, 청주 150;
조리방법: 굽기

3. 오색볶음면
요리종류 : 일품
재료 : 재료 양파(40g), 청경채(30g), 두 가지 색 파프리카(각 15g씩), 쪽파(15g), 라면(60g), 채 썬 돼지고기(30g), 양념 다진 마늘(7g), 고추기름(5g), 설탕(4g), 두반장(8g), 라면스프(3g)
조리방법: 볶기

다른 메뉴 추천을 원하십니까? (네 또는 아니오 입력) 네

4. 돼지수육맑은전골
요리종류 : 국&찌개
재료 : 통삼겹 250g, 된장 10g, 통후추 2g, 다진대파 5g, 팽이버섯 10g, 당근 20g, 배추 30g, 청경채 20g, 애호박 25g, 숙주 25g, 청양고추 20g, 옥수 : 천연다시다 15g, 물 300g, 다진마늘 5g, i
조리방법: 끓이기

실행 중(51초) Cell > recommend_service() > raw_input() > _input_request() > recv() > recv_multipart()
```

다른 메뉴 추천을 원할 시, 추가적으로 다른 메뉴들이 잘 도출됨을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 4)

```
메뉴 추천
식사 메뉴를 고르십니까, 후식을 고르십니까? 후식
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시시오. 여러 재료를 입력할 땐 ,(침표)로 구분하여 주세요. (없을 시 공백) 아몬드
채식주의자이십니까? (네 또는 아니오 입력) 아니오
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:63: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

1. 칼들 들썩 고구마죽
요리종류 : 후식
재료 : 고구마 100g(2/3개), 설탕 2g(1/3작은술), 참쌀가루 3g(2/3작은술),물 200ml(1컵), 잣 8g(8알)
조리방법: 끓이기
레시피:
1. 고구마는 깨끗이 씻어서 껍질을 벗기고 4cm 정도로 잘라준다.
2. 찜기에 고구마를 넣고 20-30분 정도 삶아 주고, 블렌더나 체를 이용하여 잘 으깨어 곱게 만든다.
3. 고구마와 물을 섞어 끓이면서 참쌀가루로 농도를 맞추고 설탕을 넣어 맛을 낸다.
4. 이것을 편에 노릇하게 볶아 다져서 고구마 죽에 섞는다. 기호에 따라 고구마를 튀겨 넣어 먹어도 좋다.

2. 적양배추 순두부 푸딩
요리종류 : 후식
재료 : 순두부 50g(1/6봉지), 우유 75g(5큰술), 생크림 25g(1½큰술), 꿀 35g(2½큰술), 판젤라틴 2, 적양배추 100g(10×5cm 8장), 꿀 35g(2½ 큰술), 판젤라틴 2, 라즈베리 8g(1큰술), 설탕 50g(5
조리방법: 끓이기
레시피:
1. 판젤라틴을 찬물에 넣어 부드러워지게 불려 물에서 건진다.
2. 불린 젤라틴을 증탕으로 액체 상태로 만든다.
3. 설탕과 물을 약한 불에서 끓이다가 라즈베리를 넣고 5분간 졸인다.
4. 순두부, 우유, 생크림, 바닐라에센스, 꿀을 섞어 블렌더를 사용하여 갈아준다.
5. 녹인 젤라틴(2)의 반을 갈아놓은 재료(4)에 넣어 잘 섞는다.
6. 적양배추는 체를 썰어 물을 자작자작하게 붓고 10분 정도 양배추가 부드러워지게 삶는다.
7. 삶은 적양배추를 체에 건져서 양배추즙을 낸다.
8. 적양배추 즙에 꿀을 넣어 섞은 후 남겨놓은 젤라틴(2)의 반을 잘 섞는다.
9. 사각 물드나 사각 등에 순두부혼합물(5)과 적양배추혼합물(8)을 각각 부어 냉장고에 넣어 굳힌다.
10. 푸딩을 꺼내어 접시에 담은 후 라즈베리 소스를 부리고 민트로 장식한다.

실행 중(21초) Cell > recommend_service() > raw_input() > _input_request() > recv() > recv_multipart()
```

후식 추천도 잘 이루어짐을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 5)

```
메뉴 추천
식사 메뉴를 고르십니까, 후식을 고르십니까? 식사!
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 아니오
드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(침표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백) 소고기
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시시오. 여러 재료를 입력할 땐 ,(침표)로 구분하여 주세요. (없을 시 공백) 복숭아
채식주의자이십니까? (네 또는 아니오 입력) 네

조건을 만족하는 메뉴가 (더이상) 없습니다.
```

원하는 식재료가 소고기인데 채식주의자라 하였으므로 조건을 만족하는 메뉴가 없는 것이 타당한 결과이다. 채식주의자 선택 시 대표적인 고기류가 잘 제외됨을 알 수 있다.

메뉴 추천 서비스 실행 결과 예시 6)

 메 뉴 추 천

식사 메뉴를 고르십니까, 후식을 고르십니까? 식사
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 아뇨
드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백) 카레, 면
특정 식재료에 알레르기 있습니까? 있다면 입력해주십시오. 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백)
채식주의자이십니까? (네 또는 아니오 입력) 아니요

1. 닭갈비볶음면

요리종류 : 일품

재료 : 재료 마늘(20g), 깻잎(2g), 쪽파(1g), 파프리카(60g), 양파(40g), 대파(10g), 방울토마토(50g), 닭가슴살(30g), 이태리시즈닝(1g), 라면(110g), 양념 카레가루(2g), 고춧가루(2g)

조리방법: 볶기

조건을 만족하는 메뉴가 (더이상) 없습니다.

 메 뉴 추 천

식사 메뉴를 고르십니까, 후식을 고르십니까? 식사
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 아니요
드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백) 고구마, 닭
특정 식재료에 알레르기가 있습니까? 있다면 입력해주세요. 여러 재료를 입력할 땐 ,(참표)로 구분하여 주세요. (없을 시 공백) 당근, 양파
채식주의자이십니까? (네 또는 아니오 입력) 아니요

1. 닭가슴살카나페

요리종류 : 반찬

재료 : 재료 고구마(100g), 우유(110g), 백김치(40g), 비트(40g), 닭가슴살(100g), 후춧가루(0.3g), 새싹(1g), 바나나식초(80g)

조리방법: 굽기

2. 복분자소스를 곁들인 닭테린

요리종류 : 반찬

재료 : 닭고기(가슴살, 200g), 시금치(30g), 브로콜리(30g), 토마토(100g), 고구마(100g), 아스파라거스(3개), 복분자(20g), 새송이버섯(20g)

조리방법: 굽기

조건을 만족하는 메뉴가 (더이상) 없습니다.

원하는 조건을 만족시키는 것이 3개 미만일 경우, 가능한 결과만을 도출하고 프로그램이 종료된다.

메뉴 추천 서비스 실행 결과 예시 7)

```

식사 메뉴를 고르십니까, 후식을 고르십니까? 식사
직접 요리를 하실 예정입니까? (네 또는 아니오 입력) 아니오
드시고 싶은 특정한 식재료가 있습니까? 여러 재료를 입력할 땐 ,(실표)로 구분하여 주세요. (e.g. 닭, 면, 버섯) (없을 시 공백) 감자, 치즈
특정 식재료에 알레르기가 있습니까? 있다면 입력해주시시오. 여러 재료를 입력할 땐 ,(실표)로 구분하여 주세요. (없을 시 공백) 토마토
채식주의자입니까? (네 또는 아니오 입력) 네
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:47: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:63: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

1. 구운고등어 & 감자수제비 파스타
요리종류 : 일품
재료 : 감자 300g(2개), 시금치 40g, 비트 20g(1/4개), 오징어먹물 2g(1/2작은술), 밀가루 200g(1컵), 달걀 120g(2개), 올리브오일 10g(2작은술), 버터10g ,다진 양파 40g(1/4개), 생크림 300g(1리터)
조리방법: 볶기

2. 산마 떡갈비와 무화과발사믹 소스
요리종류 : 반찬
재료 : 감자 300g(2개), 병가루 100g(1/2컵), 달걀 50g(1개), 파마산 치즈가루 15g(1큰술), 버터 7g(1작은술), 파슬리 5g(1작은술), 밀가루 5g(1작은술), 감자 50g(1/3개), 겨자가루 2g(1/2작은술), 식용유 10g
조리방법: 굽기

3. 가지말이샐러드
요리종류 : 반찬
재료 : 가지 250g, 감자 50g, 미니파프리카 30g, 양파 40g, 마늘튀김 15g, 소스 : 홀그레인머스터드 15g, 올리브오일 30g, 다진마늘 15g, 설탕 5g, 올리브고당 15g, 레드와인 30g, 꿀들이체소 : 어린체소 10g
조리방법: 굽기

다른 메뉴 추천을 원하십니까? (네 또는 아니오 입력) 네
```

4. 비트 단호박 파스타
요리종류 : 일품
재료 : 삶은 스파게티(80g), 감자(10g), 파르메산 치즈가루(약간), 어민일채소(약간), 비트(40g), 올리브유(30g), 스리라차소스(5g), 마늘(10g), 미니 단호박(70g), 생크림(100g), 설탕(3g)
조리방법: 기타

5. 앙송이버섯 달걀 크레페
요리종류 : 반찬
재료 : 앙송이버섯 12.8g, 달걀 11g, 에벤탈치즈 20g, 메밀가루 20g, 달걀 22g, 밀가루 20g, 우유 80g, 설탕 10g, 양파 8g, 파슬리다진것 0.4g, 감자 20g, 로즈마리다진것 0.2g, 마늘다진것 1.2g,
조리방법: 굽기

조건을 만족하는 메뉴가 더이상 없습니다.

사용자가 추가 추천을 받겠다고 입력하였지만 더이상 추천할 메뉴가 없을 때도, 가능한 메뉴를 추천한 후 프로그램이 종료된다.

4.2) 음식 수식어 분석 결과

간단	8920
맛있	7654
매콤	2340
시원	2122
고소	1902
쉽	1824
달콤	1749
얼큰	1098
아삭	1010
바삭	963
담백	926
가득	870
구수	828
들쭉	802
새콤달콤	800
칼칼	798
간편	740
상큼	735
쫄깃	691
든든	690
독막	655
향긋	637
달달	585
부드럽	568
촉촉	553

1) 가장 많이 사용된 표현

수식어 분석 결과, 왼쪽 Figure 1의 25개 표현이 가장 많이 사용된 표현임을 알 수 있었다. 음식의 맛 자체를 표현하기보다 음식 요리 과정을 수식하는 표현인 ‘간단’이 가장 많이 사용되었다는 것은 예상하지 못했던 놀라운 결과였다. 하지만 이는 데이터가 ‘레시피 제목’ 데이터라는 점에서 충분히 받아들일 수 있는 분석 결과라 본다. 직접 요리를 하는 사람들은 간단하게 만들어낼 수 있는 요리를 선호할 가능성이 크고, 이들의 이목을 끌어야 하는 레시피 글은 ‘간단’, ‘쉽’, ‘간편’과 같은 수식어를 많이 활용할 충분한 이유가 있다.

이외에 순위권에 이름을 올리고 있는 수식어들을 살펴보면, 매콤, 고소, 달콤과 같은 맛 표현 외에도 아삭, 바삭, 쫄깃, 부드럽, 촉촉과 같이 음식의 질감이나 촉감과 관련된 표현도 많이 사용됨을 알 수 있다.

Figure 1 가장 많이 사용된 수식어

또한 빈도 1, 2위와 3위 사이의 격차가 크다는 점도 눈여겨볼 만하다.

2) 글 제목에 쓰인 수식어의 빈도와 글 조회수와의 관계

modifier	
각각각각	326699.000000
자작자작	202358.666667
그득그득	166462.000000
톡톡	160639.000000
아작	148845.200000
설렁설렁	133485.000000
답답	130743.000000
찰싹	120768.000000
침빙	111982.000000
말갈	105817.000000
진부	99273.000000
따뜻	95936.000000
고프	94374.000000
엄청나	90595.000000
일찍	89662.000000
탕탕탕	86817.000000
그그	83160.000000

Figure 2 수식어별 평균 조회수

알아볼 수 있다. 사람들이 많이 쓰는 음식 수식어를 제목에 포함하는 것이 높은 조회수를 보장하지 않으며, 반대로 높은 조회수를 갖는 글의 제목에 사용된 수식어가 가장 흔히 사용되는 수식어라고 단언할 수 없다.

수식어 별로 (그 수식어가 사용된 제목을 가지는 글의) 조회수 평균을 구하여 이 둘 간의 관계가 있는지 알아보고자 하였다. Figure 2는 조회수가 높은 순으로 수식어를 정렬한 것이다. 이 결과만 봐서는 수식어의 사용 빈도와 조회수 사이 유의미한 영향은 없어 보인다. 그러나 Figure 2에는 음식을 수식하지 않는 수식어도 많이 보이므로 추가 분석이 필요하다.

Figure 3은 1)에서 구한 가장 많이 사용된 표현들의 조회수와 빈도를 한 그래프에 나타낸 것이다. 빨간 선 그래프가 조회수를 나타내고 막대 그래프가 빈도수를 나타낸다. y축은 선 그래프에 대하여 조회수를 나타내고 막대 그래프에 대하여 빈도수를 나타낸다. 그래프를 통해 우리는 수

식어의 빈도와 조회수가 특별한 상관관계를 갖지 않는다는 것을 한눈에

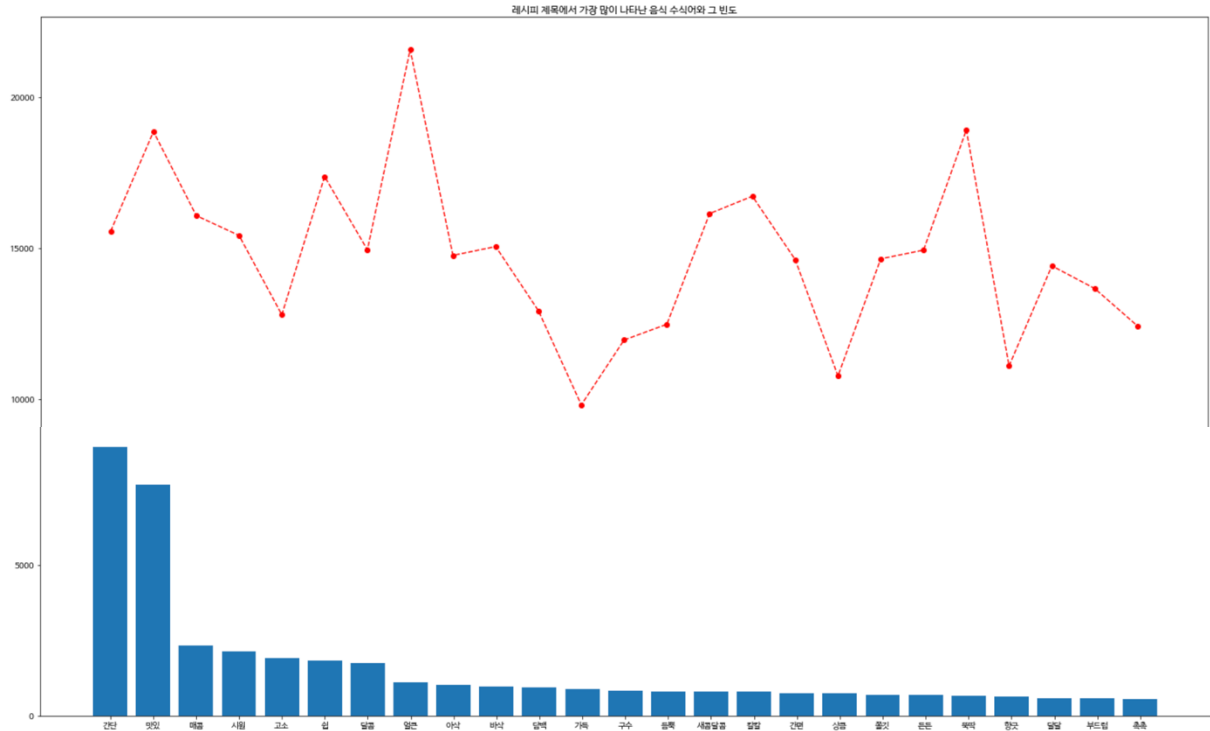


Figure 3 가장 많이 사용되는 음식 수식어 25개와, 그 수식어가 제목 내에 사용된 글의 평균 조회수

3) 연도별 수식어 사용 차이

데이터는 2007년부터 2020년까지 14년에 걸친 장기간의 데이터로, 시간에 따라 변화하는 언어 사용을 드러낼 수 있을 것이라 생각하였다. 각 연도별 사용빈도가 높은 수식어 5가지를 빈도 순으로 도출한 것은 다음과 같다.

<pre> * year 2007 * count 71 달콤 9 매콤 7 가득 5 고소 4 맛있 4 Name: modifier, dtype: int64 </pre>	<pre> * year 2010 * count 1671 간단 117 매콤 89 좋 85 맛있 83 달콤 82 Name: modifier, dtype: int64 </pre>	<pre> * year 2013 * count 1253 맛있 122 간단 68 좋 64 고소 63 달콤 56 Name: modifier, dtype: int64 </pre>	<pre> * year 2016 * count 10059 간단 1043 맛있 1011 좋 565 매콤 336 시원 327 Name: modifier, dtype: int64 </pre>
<pre> * year 2008 * count 771 달콤 47 좋 47 간단 36 매콤 36 고소 30 Name: modifier, dtype: int64 </pre>	<pre> * year 2011 * count 954 달콤 64 좋 59 간단 59 고소 51 매콤 43 Name: modifier, dtype: int64 </pre>	<pre> * year 2014 * count 941 맛있 103 좋 87 간단 75 고소 38 시원 36 Name: modifier, dtype: int64 </pre>	<pre> * year 2017 * count 10392 간단 1313 맛있 994 좋 548 매콤 285 시원 281 Name: modifier, dtype: int64 </pre>
<pre> * year 2009 * count 1370 좋 81 매콤 79 맛있 64 고소 61 달콤 61 Name: modifier, dtype: int64 </pre>	<pre> * year 2012 * count 1067 맛있 103 달콤 64 고소 62 좋 59 간단 58 Name: modifier, dtype: int64 </pre>	<pre> * year 2015 * count 7356 간단 829 맛있 777 좋 479 매콤 240 시원 229 Name: modifier, dtype: int64 </pre>	<pre> * year 2018 * count 13104 간단 1561 맛있 1150 좋 795 매콤 343 시원 320 Name: modifier, dtype: int64 </pre>
<pre> * year 2019 * count 12141 간단 1671 맛있 1245 좋 682 매콤 343 시원 302 Name: modifier, dtype: int64 </pre>	<pre> * year 2020 * count 16644 간단 2038 맛있 1929 좋 816 쉽 506 시원 427 Name: modifier, dtype: int64 </pre>		

연도별 절대적 수치의 차이는 있지만, 각 연도별로 수식어의 순위만 두고 본다면 두 가지 흥미로

운 특징이 관찰된다.

첫 번째는 ‘간단’의 순위에 관한 것이다. 2007년과 2009년만 하더라도 ‘간단’이란 수식어는 순위에서 찾아볼 수 없다. 2008년, 2010년, 그리고 2012년의 데이터만 보더라도 해당 수식어의 사용 빈도는 변동을 거듭한다. 그러나 2015년부터 ‘간단’은 사용빈도 1위를 놓친 적이 없다. 시간이 흐르면 흐를수록 ‘간단’과 빈도 순위 3위 이하의 수식어들 간 격차가 커지는 것도 흥미롭다. 이는 사람들이 점점 더 간단하게 해먹을 수 있는 요리를 선호한다는 점을 드러낼 수 있을 것이다. 2020년에 수식어 ‘쉽’의 등장은 이를 더욱 뒷받침한다.

두 번째는 ‘달콤’의 순위에 관한 것이다. ‘달콤’은 2007년부터 2013년까지 꾸준히 빈도 순위권에 올라있다. 사용 빈도 1위를 차지한 적도 있다. 그러나 2014년부터는 어째선지 순위권에서 찾아볼 수 없다. 이는 사람들의 ‘달콤’한 맛에 대한 선호도가 전반적으로 낮아졌음을 시사할 수도 있을 것이다.

5. 의의 및 한계

본 프로젝트는 레시피 데이터를 분석하여 메뉴 추천 서비스와 음식 수식어 분석 양상 결과를 제공하였다. ‘메뉴추천 서비스’는 식사 메뉴를 고민하는 상황에 있는 사람들에게 도움을 줄 수 있을 것이다. 사람들은 원하는 식재료와 제외하고 싶은 식재료를 고려하여 메뉴에 대한 새로운 선택지를 제공받을 수 있고, 추천 받은 음식을 만들 수 있는 레시피까지 알 수 있다. 이는 실생활의 효용 증진에 기여할 수 있으리라 기대한다.

음식 수식어 분석을 통해서도 사람들의 실제 언어 사용 양상을 분석하였다. 사람들은 음식을 표현할 때 맛과 관련된 표현 뿐 아니라 음식의 질감을 표현할 수 있는 수식어도 다수 활용함을 알 수 있었다. 또한 음식을 만드는 과정에서는 ‘간단함’과 ‘간편함’이 중요한 요소로 부상하였다는 점도 흥미로웠다. 나아가 수식어와 빈도와 그 수식어가 사용된 제목을 가지는 글 조회수는 유의미한 상관관계를 가지지 못하였다.

프로젝트 분석에서 다양한 요소를 고려하여 언어 데이터를 정확히 분석하고자 노력하였다. 그러나 아직 해결하지 못한 많은 문제들이 남아있다. 본 프로젝트에서는 ‘매콤’이 명사로 분류되었을 때의 빈도수를 반영하지 못하였는데, 이는 명사 중에서 수식어로 활용될 수 있는 것들을 분류하기 어렵다는 점에서 기인한다. 또한, 채식주의자 관련 데이터 필터링을 진행할 때 ‘소’고기를 제외하고 싶었는데 ‘소’라는 글자만으로는 그것이 소고기의 ‘소’인지 소금의 ‘소’인지, 어떤 의미를 갖고 있는 것인지 알 수 없어 조금 더 좁은 범위인 ‘소고기’ 혹은 ‘쇠고기’만을 필터링의 대상으로 삼았다. 이때문에 분석의 정확성이 떨어졌을 것이다. 마지막으로, 포함 혹은 제외하고 싶은 식재료로 소고기나 계란을 입력하였을 때 이와 동의어인 쇠고기와 달걀에도 동일한 처리를 할 수 있었다면 더 좋았을 것이다. 이러한 한계점을 극복한다면 성능을 개선하고 더욱 정확한 결과를 도출해낼 수 있을 것이라 기대한다.