터 

**SK네트웍스 Family AI 과정 10기  
 모델링 및 평가 수집된 데이터 및 전처리 문서**



| **산출물 단계** | 모델링 및 평가 |
| --- | --- |
| **평가 산출물** | 수집된 데이터 및 전처리 문서 |
| **제출 일자** | 2025-09-12 |
| **깃허브 경로** | https://github.com/skn-ai14-250409/SKN14-Final-6Team-AI |
| **작성 팀원** | 서은선 |

1. **데이터 수집 개요**

데이터 수집 기간: 2025-08-27~2025-08-29

수집 방식: 공식 페이지에서 직접 발췌 후 JSON 파일로 정리

1.1. 수집 목적

수집 목적: 신선식품 플랫폼에서 고객 문의에 정확하고 일관된 챗봇 응답을 제공하기 위해, 마켓컬리 FAQ와 이용약관을 수집하여 rag 지식베이스로 활용하기 위한 목적으로 수집합니다.

1.2 데이터 출처

| **데이터명** | **출처** | **수집 방식** | **형식** |
| --- | --- | --- | --- |
| 마켓컬리 FAQ | 마켓컬리 홈페이지 | 직접 발췌 | JSON |
| 마켓컬리 이용약관 | 마켓컬리 페이지 | 직접 발췌 | JSON |

1.3 수집 대상 및 범위

* 대상: 마켓컬리 faq 및 이용약관
* 총 수집 건수:
  + 마켓컬리 FAQ: 약 117건
  + 마켓컬리 이용약관: 47건
* 개인정보 포함 여부: 아니오

**2.저장 방식**

* 저장 경로:
  + 마켓컬리 FAQ: /data/faq.json
  + 마켓컬리 이용약관: /data/terms\_of\_use.json
* 저장 형식: JSON

**3. 데이터 전처리 절차**

[마켓컬리 FAQ]

3.1 텍스트 정규화:

* 텍스트 정규화::  
  + 고유명사 치환: “마켓컬리” -> “Qook”
  + 특수문자 치환: “▣ Qook패스” → “■ Qook패스”
* 사용 라이브러리: re
* 전처리 전/후 비교
  + 전처리 전

{

"category":"배송",

"question":"낮배송 지역인데 배송이 언제 되나요?",

"answer":"■ 낮배송(택배)지역 배송일정 안내\n낮배송의 경우, 당일 밤 10시까지 주문하시면 다음날 밤 12시전까지 배송해드리고 있습니다.\n단, 월~토요일까지 배송하나 일요일, 공휴일은 휴무로 주 6일 배송하는 점 참고 바랍니다.\n[참고]\n▣ 일요일 배송 불가 → 토요일 주문 불가\n▣ 공휴일 배송 불가 → 공휴일 전일 주문 불가\n\* 택배사 사정으로 인해 배송 불가인 경우 사전에 공지사항 또는 마켓컬리홈 팝업창을 통해 안내 드립니다.\n\* 배송 과정 중 기상 악화 및 도로 교통 상황에 따라 부득이하게 지연 배송이 발생될 수 있습니다.\n▣ 대구/부산/울산 지역은 특정 시간대부터 택배배송 주문으로 진행 됩니다. 대구 : 20시~22시, 부산/울산 : 18시~22시"

}

* + 전처리 후

{

"category":"배송",

"question":"낮배송 지역인데 배송이 언제 되나요?",

"answer":"■ 낮배송(택배)지역 배송일정 안내\n낮배송의 경우, 당일 밤 10시까지 주문하시면 다음날 밤 12시전까지 배송해드리고 있습니다.\n단, 월~토요일까지 배송하나 일요일, 공휴일은 휴무로 주 6일 배송하는 점 참고 바랍니다.\n[참고]\n■ 일요일 배송 불가 → 토요일 주문 불가\n■ 공휴일 배송 불가 → 공휴일 전일 주문 불가\n\* 택배사 사정으로 인해 배송 불가인 경우 사전에 공지사항 또는 Qook홈 팝업창을 통해 안내 드립니다.\n\* 배송 과정 중 기상 악화 및 도로 교통 상황에 따라 부득이하게 지연 배송이 발생될 수 있습니다.\n▣ 대구/부산/울산 지역은 특정 시간대부터 택배배송 주문으로 진행 됩니다. 대구 : 20시~22시, 부산/울산 : 18시~22시"

}

[마켓컬리 이용약관]

3.2 텍스트 정규화:

* 텍스트 정규화::  
  + 고유명사 치환: “마켓컬리” -> “Qook”
* 사용 라이브러리: re
* 전처리 전/후 비교
  + 전처리 전

{

"article\_number": "마켓컬리 이용약관 제1조",

"title": "목적",

"content": [

"① 이 약관은 주식회사 마켓컬리(이하 주식회사 마켓컬리 또는 주식회사 마켓컬리가 운영하는 인터넷사이트를 “마켓컬리”라 함)가 운영하는 온라인쇼핑몰에서 제공하는 전자상거래 관련 서비스(이하 “서비스”라 함)를 이용함에",

"② 있어 마켓컬리와 이용자의 권리, 의무 및 책임사항을 규정함을 목적으로 합니다.",

"③ \*PC통신, 스마트폰 앱, 무선 등을 이용하는 전자상거래에 대해서도 그 성질에 반하지 않는 한 이 약관을 준용합니다."

]

}

* + 전처리 후

{

"article\_number": "Qook 이용약관 제1조",

"title": "목적",

"content": [

"① 이 약관은 주식회사 Qook(이하 주식회사 Qook 또는 주식회사 Qook가 운영하는 인터넷사이트를 “Qook”라 함)가 운영하는 온라인쇼핑몰에서 제공하는 전자상거래 관련 서비스(이하 “서비스”라 함)를 이용함에",

"② 있어 Qook와 이용자의 권리, 의무 및 책임사항을 규정함을 목적으로 합니다.",

"③ \*PC통신, 스마트폰 앱, 무선 등을 이용하는 전자상거래에 대해서도 그 성질에 반하지 않는 한 이 약관을 준용합니다."

]

}

3.3 구조화

* 구조화 방식:
* 최상위 조항(①,②,…) → 새로운 Document 시작
* 숫자 항목(1,2,3…) → / 기호로 구분
* 한글 소분류(가,나,다…) → // 기호로 구분
* 사용 라이브러리: re
* 전처리 전/후 비교
  + 전처리 전

"article\_number": "마켓컬리 이용약관 제4조",

"title": "서비스의 제공 및 변경",

"content": [

"① 마켓컬리는 다음과 같은 업무를 수행합니다.",

"1. 전자상거래 서비스(통신판매중개 서비스 포함) 및 이에 수반되는 기타 서비스",

"2. 기타 마켓컬리가 정하는 업무",

"② 마켓컬리는 상품 등의 품절 또는 기술적 사양의 변경 등 사유가 발생하는 경우에는 장차 체결되는 계약에 의해 제공할 상품 등의 내용을 변경할 수 있습니다. 이 경우에는 변경된 상품 등의 내용 및 제공일자를 명시하여 현재의 상품 등의 내용을 게시한 곳에 즉시 공지합니다.",

"③ 마켓컬리k가 제공하기로 이용자와 계약을 체결한 서비스의 내용을 상품 등의 품절 또는 기술적 사양의 변경 등의 사유로 변경할 경우에는 그 사유를 이용자에게 즉시 통지합니다."

]

},

* + 전처리 후

{

"article\_number": "Qook 이용약관 제4조",

"title": "서비스의 제공 및 변경",

"content": [

"① Qook는 다음과 같은 업무를 수행합니다./1. 전자상거래 서비스(통신판매중개 서비스 포함) 및 이에 수반되는 기타 서비스/2. 기타 Qook가 정하는 업무",

"② Qook는 상품 등의 품절 또는 기술적 사양의 변경 등 사유가 발생하는 경우에는 장차 체결되는 계약에 의해 제공할 상품 등의 내용을 변경할 수 있습니다. 이 경우에는 변경된 상품 등의 내용 및 제공일자를 명시하여 현재의 상품 등의 내용을 게시한 곳에 즉시 공지합니다.",

"③ Qook가 제공하기로 이용자와 계약을 체결한 서비스의 내용을 상품 등의 품절 또는 기술적 사양의 변경 등의 사유로 변경할 경우에는 그 사유를 이용자에게 즉시 통지합니다."

]

},

3.4 전처리 절차 요약

[마켓컬리 FAQ]

1. 데이터 수집
2. 정규화(텍스트 전처리)
3. 정규화(고유명사 치환)
4. 저장

[마켓컬리 이용약관]

1. 데이터 수집
2. 정규화(고유명사 치환)
3. 구조화
4. 저장

**4. 부록**

* 수집 데이터 샘플
  + 마켓컬리 FAQ

{

"category":"취소/교환/환불",

"question":"교환(반품) 어떻게 진행 되나요?",

"answer": "■ 교환(반품) 진행 방법\n- 받으신 상품을 교환(반품)하실 경우, 교환 사유와 문제가 발생한 부분을 확인할 수 있는 사진과 함께 고객센터로 문의 바랍니다.\n- 상담사를 통한 상담 이후, 교환(반품) 절차가 진행됩니다.\n(PC) 홈페이지 상단 고객센터 > 1:1 문의 또는 홈페이지 하단 1:1 문의\n(모바일) 마이Qook > 1:1 문의 또는 마이Qook > 고객센터 > 1:1 문의 또는 카카오톡 문의"

* + 마켓컬리 이용약관

{

"article\_number": "Qook 이용약관 제4조",

"title": "서비스의 제공 및 변경",

"content": [

"① Qook는 다음과 같은 업무를 수행합니다.",

"1. 전자상거래 서비스(통신판매중개 서비스 포함) 및 이에 수반되는 기타 서비스",

"2. 기타 Qook가 정하는 업무",

"② Qook는 상품 등의 품절 또는 기술적 사양의 변경 등 사유가 발생하는 경우에는 장차 체결되는 계약에 의해 제공할 상품 등의 내용을 변경할 수 있습니다. 이 경우에는 변경된 상품 등의 내용 및 제공일자를 명시하여 현재의 상품 등의 내용을 게시한 곳에 즉시 공지합니다.",

"③ Qook가 제공하기로 이용자와 계약을 체결한 서비스의 내용을 상품 등의 품절 또는 기술적 사양의 변경 등의 사유로 변경할 경우에는 그 사유를 이용자에게 즉시 통지합니다."

]

}

* 전처리 코드
  + 마켓컬리 FAQ

text = re.sub(r"[■▣▶※◆◇●]", "■", text)

return re.sub(r"(마켓컬리|컬리)", "Qook", text)

* + 마켓컬리 이용약관

# ①,②,… 최상위 항목

if re.match(r'^[①-⑳]', line):

if buffer and current\_article:

docs.append(Document(

page\_content=f"{current\_article} " + " ".join(buffer),

metadata={

"article": item["article\_number"],

"title": item["title"],

"type": "terms"

} ))

elif re.match(r'^\d+', line):

buffer.append("/" + line)

elif re.match(r'^[가-힣]', line):

if buffer:

buffer[-1] += " //" + line

else:

buffer.append("//" + line)

* 참조 문헌
  + 마켓컬리 FAQ: https://www.kurly.com/m2/service/faq.php
  + 마켓컬리 이용약관: https://www.kurly.com/user-terms/agreement