

농산물 숏폼 커머스 환경에서 RAG 기반 소비자 질의응답 시스템 구축 방안

문지원*, 조윤찬**, 장승현**, 김승현*, 최제호***, 임용표****, 진주완**
*건양대학교 방재보안학과, **건양대학교 재난안전소방학과, ***건양대학교 스마트농산업학과,
****건양대학교 스마트팜학부
e-mail:25609507@konyang.ac.kr

Design and Implementation of a RAG-based Consumer Question Answering System for Short-form Agricultural Commerce

Jiwon-Moon*, Yun-Chan Jo**, Seunghyeon Jang**, SeungHan Kim*, Je-Ho Choi***,
Yongpyo Im****, Juan Jin**
*Dept. of Disaster Prevention and Security, **Dept. of Disaster Safety & Fire Fighting,
ing, ***Dept. of Smart Agricultural Technology and Innovation, ****Dept. of Smart Farm, Konyang University

요약

본 연구는 농산물 숏폼 커머스 환경에서 소비자의 상품 정보 확인 요구에 대응하기 위한 RAG 기반 소비자 질의응답 시스템 구축 방안을 제안한다. 본 연구에서 농산물 숏폼 커머스를 지역 농가의 농산물, 농가 체험, 지역 콘텐츠를 숏폼 영상과 위치 기반 플랫폼 구조를 통해 노출하고, 소비자가 영상 시청 이후 상품 구매 또는 체험 예약으로 전환할 수 있도록 지원하는 커머스 모델을 의미한다. 농산물과 농가 체험 상품은 원산지, 생산자, 수확 시기, 재배 방식, 체험 조건, 배송 및 환불 정책 등 구매 전 확인해야 할 정보가 많지만, 기존 온라인 상품 상세페이지와 숏폼 콘텐츠는 제한된 정보를 일방향적으로 제공하는 한계가 있다. 이에 본 연구는 농가별 상품정보, 생산정보, 체험정보, 후기정보, 유통정보, 정책정보를 지식베이스화하고, 소비자 질의에 대해 검색 증강 생성 방식으로 근거 기반 응답을 제공하는 시스템 구조를 제안한다. 제안 시스템은 데이터 수집, 문서 전처리, 임베딩, 벡터 검색, 재순위화, 응답 생성, 출처 표시 및 응답 검증 모듈로 구성된다. 또한 검색 정확도, 응답 충실성, 응답 관련성, 환각 발생률, 출처 제시율, 응답 지연시간을 중심으로 시스템 평가 방안을 제시한다. 본 연구는 RAG를 농업 생산자 대상 지식 상담 도구가 아니라, 농산물 커머스에서 소비자의 정보 확인과 신뢰 형성을 지원하는 질의응답 시스템으로 확장했다는 점에서 의의가 있다.

1. 서론

최근 온라인 커머스는 상품 상세페이지 중심에서 숏폼 영상, 라이브커머스, 소셜미디어 기반 판매 방식으로 확장되고 있다. 통계청에 따르면 2024년 10월 온라인쇼핑 거래액 중 모바일쇼핑 비중은 76.0%였으며, 음식료품과 농축수산물 거래액도 전년 동월 대비 각각 17.5%, 16.3% 증가하였다[1]. 이는 농산물 유통에서도 모바일 기반 탐색과 구매가 중요한 소비 접점이 되고 있음을 보여준다.

숏폼 콘텐츠는 생산 현장, 수확 장면, 생산자 스토리 등을 짧은 시간 안에 전달할 수 있어 농산물 홍보에 적합하다. 그러나 농산물과 농가 체험 상품은 원산지, 생산자, 수확일, 재배 방식, 체험 조건, 배송 및 환불 기준 등 구매 전 확인해야 할 정보가 많다. 기존 연구에서도 숏폼 영상 콘텐츠가 소비자 신뢰와 구매의도에 영향을 미치는 요인임이 확인되었으나[2], 농산물 온라인 거래에서는 정보 제공 측면의 개선 필요성도 지적되고 있다[3].

본 논문에서 제안하는 농산물 숏폼 커머스는 지역 농가의 농산

물, 체험 프로그램, 지역 콘텐츠를 숏폼 영상과 위치 기반 플랫폼을 통해 제공하고, 소비자가 상품 구매 또는 농가 방문으로 전환할 수 있도록 지원하는 서비스 모델이다. 이러한 플랫폼에서는 소비자 질의에 대해 최신 상품정보와 정책정보에 근거한 응답을 제공하는 기능이 중요하다.

이에 본 논문에서는 농산물 숏폼 커머스 환경에서 RAG 기반 소비자 질의응답 시스템 구축 방안을 제안한다. RAG는 외부 지식베이스 검색 결과를 언어모델의 응답 생성 과정에 결합하여 사실 기반 응답을 제공하는 방식이다[4]. 본 논문은 농가별 상품정보, 생산정보, 체험정보, 후기정보, 유통정보, 정책정보를 지식베이스화하고, 소비자 질의에 대해 근거 기반 응답을 제공하는 시스템 구조와 처리 절차를 제시하는 것을 목적으로 한다.

2. 이론적 배경

2.1 농산물 온라인 커머스와 정보 제공의 중요성

농산물 온라인 커머스에서는 상품 품질, 정보 품질, 서비스 품질

질이 소비자의 신뢰 형성에 중요한 요인으로 작용한다. 농산물 전자상거래 특성이 지속사용의도에 미치는 영향을 분석한 연구에서는 농산물 품질, 시스템 품질, 정보 품질, 서비스 품질 등이 신뢰 및 지속사용의도와 관련된 주요 요인으로 제시되었다 [5]. 이는 농산물 플랫폼에서 단순한 상품 노출뿐 아니라, 소비자가 판단 가능한 수준의 정보 제공 구조가 필요함을 의미한다.

농산물은 생산 시점, 생산 지역, 유통 과정, 보관 조건에 따라 품질이 달라질 수 있다. 따라서 농산물 커머스 시스템은 상품명과 가격뿐 아니라 수확일, 포장일, 출고일, 생산자 정보, 인증 여부, 배송 조건 등 소비자가 신뢰 판단에 활용할 수 있는 데이터를 구조화해야 한다. 특히 슛폼 콘텐츠는 시각적 몰입을 제공할 수 있지만, 상세한 검증 정보를 모두 담기 어렵기 때문에 영상 이후 소비자 질의에 대응하는 보완 시스템이 요구된다.

2.2 식품 이력추적과 데이터 표준

식품 및 농산물 분야에서 정보 신뢰성을 확보하기 위해서는 생산·가공·유통·판매 단계의 데이터를 추적 가능하게 관리할 필요가 있다. ISO 22005:2007은 식품 및 사료 공급망에서 이력추적 시스템의 설계와 구현을 위한 일반 원칙과 기본 요구사항을 제시한다 [6]. 이 표준은 특정 조직이 식품 공급망의 어느 단계에 있는 적용 가능하도록 설계되었으며, 농산물 커머스 플랫폼의 데이터 관리 구조를 설계할 때 참고할 수 있다.

또한 GS1 EPCIS는 공급망에서 발생하는 이벤트 데이터를 공통 언어로 공유하기 위한 표준으로, 상품이 언제, 어디서, 어떤 상태로 이동했는지를 표현하는 데 활용될 수 있다 [7]. 농산물 슛폼 커머스에서 이러한 개념을 적용하면 수확, 포장, 출고, 배송, 체험 예약과 같은 이벤트를 데이터화하고, 소비자 질의응답 시스템의 검색 대상 지식으로 활용할 수 있다.

2.3 RAG 기반 질의응답 기술 분석

RAG는 외부 문서 검색과 생성형 언어모델을 결합한 질의응답 방식이다. Lewis et al.은 RAG를 사전학습 모델의 매개변수 지식과 비매개변수 메모리인 외부 문서 색인을 결합하는 방식으로 제안하였다[4]. 이 구조는 대규모 언어모델이 자체 학습 지식에만 의존할 때 발생할 수 있는 최신성 부족, 출처 불명확성, 사실 오류 문제를 완화하는 데 활용될 수 있다.

RAG 시스템의 핵심은 검색기와 생성기의 결합이다. 검색기는 사용자 질의와 관련된 문서를 지식베이스에서 찾고, 생성기는 검색된 문서를 바탕으로 자연어 응답을 생성한다. 문장 임베딩 기반 검색은 문장 간 의미적 유사도를 계산하여 관련 문서를 찾는 데 활용될 수 있으며, Sentence-BERT는 문장 임베딩을 효율적으로 생성하고 코사인 유사도 기반 비교가 가능하도록 제안된 대표적 방법이다[8].

검색 품질은 RAG 응답 품질에 직접적인 영향을 미친다.

Dense Passage Retrieval 연구는 조밀 벡터 표현을 활용한 문서 검색 방식이 오픈도메인 질의응답에서 효과적으로 활용될 수 있음을 보였다[9]. 농산물 커머스 환경에서도 소비자 질문과 관련된 상품정보, 생산정보, 체험정보, 정책정보를 정확히 검색하는 것이 응답 신뢰성 확보의 핵심이다.

2.4 RAG 응답 검증과 평가

RAG 시스템은 외부 지식을 활용한다는 장점이 있지만, 검색된 문서가 부정확하거나 질문과 관련성이 낮을 경우 응답 품질이 저하될 수 있다. Self-RAG는 검색 필요성 판단, 생성, 자기비판 과정을 통해 언어모델이 검색과 응답 품질을 스스로 평가하도록 설계된 접근이다[10]. 또한 Corrective RAG는 검색 결과의 품질을 평가하고, 검색 결과가 부적절할 경우 보정 과정을 수행하여 RAG의 견고성을 높이는 방법을 제안하였다[11].

RAG 시스템 평가를 위해서는 단순한 응답 유창성보다 검색 문맥의 적절성, 응답의 근거 일치성, 답변 관련성 등을 평가해야 한다. RAGAS는 RAG 파이프라인을 대상으로 faithfulness, answer relevancy, context precision, context recall 등을 평가할 수 있는 프레임워크로 제안되었다[12]. 이에 농산물 슛폼 커머스의 소비자 질의응답 시스템 역시 검색 정확도와 생성 응답의 근거 일치성을 중심으로 평가할 필요가 있다.

3. 제안 시스템 개요

3.1 시스템 구축 목표

본 연구에서 제안하는 시스템의 목표는 농산물 슛폼 커머스 플랫폼에서 소비자가 상품 또는 체험 정보를 질문했을 때, 플랫폼 내부의 검증된 데이터를 검색하여 근거 기반 응답을 제공하는 것이다. 제안 시스템은 단순히 자연스러운 답변을 생성하는 챗봇이 아니라, 소비자 질의에 대해 관련 데이터의 출처와 기준일을 함께 제시하는 정보검증형 질의응답 시스템을 지향한다. 제안 시스템의 주요 구축 목표는 다음과 같다.

첫째, 농가별로 분산된 상품정보, 생산정보, 체험정보, 후기정보, 유통정보, 정책정보를 통합 지식베이스로 구축한다.

둘째, 소비자 질의를 의도별로 분류하고, 관련성이 높은 문서 또는 데이터를 검색한다.

셋째, 검색된 근거를 바탕으로 자연어 응답을 생성하되, 근거 문서에 없는 내용은 생성하지 않도록 제한한다.

넷째, 응답과 함께 출처, 기준일, 확인 필요 여부를 제공하여 소비자가 정보의 신뢰성을 판단할 수 있도록 한다.

다섯째, 검색 정확도와 응답 충실성을 기준으로 시스템 성능을 평가한다.

3.2 전체 시스템 구조

제안 시스템은 아래 표 1과 같이 크게 데이터 계층, 검색 계층, 생성 계층, 사용자 인터페이스 계층으로 구성된다.

[표 1] RAG 기반 소비자 질의응답 봇 시스템 구조

계층	구성 요소	주요 기능
데이터 계층	상품·농가·생산·체험·후기·정책 DB	농산물 및 농가 체험 관련 원천 데이터 저장
전처리 계층	문서 정제, 토크닝, 메타데이터 부여	검색 가능한 문서 단위로 변환
검색 계층	임베딩 모델, 벡터 DB, 키워드 검색, 재순위화	소비자 질의와 관련된 근거 검색
생성 계층	LLM, 프롬프트 템플릿, 응답 생성기	검색 문맥 기반 응답 생성
검증 계층	출처 표시, 근거 일치성 검사, 불확실성 처리	응답 신뢰성 보완
UI 계층	숏폼 화면 상품 상세, 챗봇, 지도 기반 농가 페이지	소비자 질의 입력 및 응답 제공

본 구조에서 숏폼 콘텐츠는 소비자의 관심을 유도하는 진입점 역할을 하며, RAG 질의응답 시스템은 숏폼 콘텐츠만으로 해결되지 않는 세부 정보 확인 요구를 처리하는 보완 기능으로 작동한다.

4. RAG 기반 소비자 질의응답 봇 데이터 구조 설계

4.1 데이터 유형 정의

본 논문에서는 농산물 숏폼 커머스에서 RAG 기반 소비자 질의응답 봇을 구현하기 위해 데이터 검색 가능한 형태로 정리되어야 한다. 플랫폼 데이터는 다음 표 2와 같이 구분한다.

[표 2] RAG 기반 소비자 질의응답 봇 데이터 유형 정의

데이터 유형	주요 항목	소비자 질의 예시
농가 기본정보	농가명, 주소, 생산자, 운영 기간, 인증 여부	“이 농가는 어디에 있나요?”
상품정보	품목, 가격, 중량, 원산지, 품종, 보관방법	“이 사과는 어떤 품종인가요?”
생산정보	수확일, 재배방식, 농약 사용 여부, 생산 수량	“이 딸기는 언제 수확했나요?”
체험정보	체험명, 예약 가능일, 인원, 소요시간, 우천 시 운영 여부	“비가 와도 체험이 가능한가요?”
유통정보	포장일, 출고일, 배송사, 예상 도착일	“오늘 주문하면 언제 도착하나요?”
정책정보	취소·환불 기준, 교환 조건	“예약 취소 시 환불되나요?”
후기정보	평점, 후기 내용, 반복 불만, 긍정 키워드	“배송 관련 불만이 많나요?”
숏폼 메타데이터	영상 제목, 태그, 등장 상품, 촬영 위치, 업로드일	“영상에 나온 상품이 이 상품 맞나요?”

4.2 메타데이터 설계

RAG 검색 품질을 높이기 위해 각 문서에는 메타데이터를 다

음 표 3과 같이 부여한다. 메타데이터는 검색 범위를 제한하고, 응답 생성 시 출처를 표시하는 데 활용된다.

[표 3] RAG 기반 소비자 질의응답 봇 메타데이터 설계

메타데이터 항목	설명
farm_id	농가 고유 식별자
product_id	상품 고유 식별자
experience_id	체험 상품 고유 식별자
document_type	상품정보, 생산정보, 정책정보, 후기정보 등 문서 유형
location	농가 또는 체험 장소
date	정보 기준일 또는 등록일
validity_period	정보 유효기간
source	데이터 출처
confidence_level	데이터 신뢰 수준
update_status	최신 정보 여부

예를 들어 “이 딸기는 언제 수확했나요?”라는 질문은 product_id, farm_id, document_type=생산정보, date 기준으로 검색될 수 있다. 반면 “환불이 가능한가요?”라는 질문은 document_type=정책정보를 우선 검색하도록 설계할 수 있다.

4.3 문서 토크닝 방식

농산물 데이터는 상품 상세페이지, 생산일지, 체험 안내문, 환불정책, 후기 등 다양한 형식으로 존재한다. 따라서 하나의 긴 문서를 그대로 임베딩하기보다 의미 단위로 분할하는 토크닝 과정이 필요하다. 아래 표 4는 RAG 기반 소비자 질의응답 봇의 토크닝 수행 기준이다.

[표 4] RAG 기반 소비자 질의응답 봇의 토크닝 수행 기준

문서 유형	토크닝 기준
상품 상세정보	상품명, 품종, 가격, 중량, 보관방법 단위
생산정보	수확일, 재배방식, 인증정보, 농약 사용 여부 단위
체험 안내	예약일, 운영시간, 인원, 준비물, 우천 기준 단위
정책 문서	배송, 교환, 환불, 취소 기준 단위
후기 데이터	개별 후기 또는 이슈별 요약 단위
숏폼 설명	영상 제목, 자막, 태그, 등장 상품 단위

5. RAG 기반 소비자 질의응답 처리 및 평가 방안

제안 시스템은 소비자 질의 입력, 의도 분류, 관련 문서 검색, 응답 생성, 출처 표시의 순서로 동작한다. 소비자가 숏폼 영상이나 상품 상세페이지에서 질문을 입력하면, 시스템은 질의가 상품정보, 생산정보, 체험정보, 배송정보, 환불정책, 후기정보 중 어디에 해당하는지 분류한다. 이후 농가별 상품정보, 생산이력, 체험 안내, 후기, 정책 문서를 검색하고, 검색된 근거를 바탕으로 응답을 생성한다.

농산물 커머스 질의는 상품명, 농가명, 지역명처럼 정확한 키워드가 중요한 경우와 “비가 오면 체험이 가능한가요?”처럼 의미 기반 검색이 필요한 경우가 함께 존재한다. 따라서 본 연구는 키

워드 검색과 벡터 검색을 결합한 하이브리드 검색 방식을 적용한다. 또한 응답에는 출처와 기준일을 함께 제시하며, 근거가 부족한 경우에는 임의로 답변하지 않고 확인 필요 항목으로 표시한다.

시스템 평가는 검색 성능과 응답 품질로 구분한다. 검색 성능은 Top-k Accuracy, Recall@k, Precision@k, MRR로 평가하고, 응답 품질은 Faithfulness, Answer Relevance, Hallucination Rate, Source Citation Rate로 평가한다. 이를 통해 일반 LLM 응답, 기본 RAG 응답, 제안 RAG 응답을 비교하여 제안 시스템의 적합성을 검증한다[12].

6. 결론

본 연구는 농산물 샵콤 커머스 환경에서 소비자 질의에 근거 기반으로 응답하는 RAG 기반 질의응답 시스템 구축 방안을 제안하였다. 제안 시스템은 농가별 상품정보, 생산정보, 체험정보, 후기정보, 유통정보, 정책정보를 지식베이스화하고, 소비자 질문에 대해 관련 근거를 검색한 뒤 응답을 생성한다.

이를 통해 소비자는 수확일, 원산지, 재배 방식, 체험 조건, 배송 및 환불 기준 등을 즉시 확인할 수 있으며, 농가는 반복적인 고객 문의를 줄일 수 있다. 또한 플랫폼은 질의응답 데이터를 활용하여 FAQ 자동 생성, 상품 상세정보 개선, 후기 요약, 추천 서비스 고도화로 확장할 수 있다.

본 연구의 의의는 RAG를 농업 생산자 대상 상담 도구가 아니라, 농산물 커머스에서 소비자의 정보 확인과 신뢰 형성을 지원하는 시스템으로 확장했다는 점에 있다. 향후 연구에서는 실제 농가 데이터를 기반으로 프로토타입을 구현하고, 검색 정확도와 응답 충실성을 실증적으로 검증할 필요가 있다.

감사의 글

이 논문은 SW중심대학사업단 Lab-Corps 프로그램의 지원을 받아 작성되었습니다.

참고문헌

- [1] 통계청, “「2024년 10월 온라인쇼핑동향」”, 2024년
- [2] Luo, C., et al. (2025). “Influence of short video content on consumers purchase intentions on social media platforms with trust as a mediator.” *Scientific Reports*, 15, 16681
- [3] 한국농촌경제연구원, “「농산물 온라인마케팅 실태와 개선과제」”, 2018년
- [4] Lewis, (2020). “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” *Advances in Neural*

- Information Processing Systems*, 33
- [5] 강지현 외, “농산물 전자상거래 특성이 지속사용의도에 미치는 영향에 관한 연구: 소셜커머스를 중심으로”, *한국경영학회 학술지*, pp.3682-3709, 2014년
- [6] ISO. (2007). ISO 22005:2007, Traceability in the feed and food chain — General principles and basic requirements for system design and implementation.
- [7] GS1. EPCIS and CBV Standards.
- [8] Reimers, N., & Gurevych, I. (2019). “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.” *Proceedings of EMNLP-IJCNLP 2019*.
- [9] Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). “Dense Passage Retrieval for Open-Domain Question Answering.” *Proceedings of EMNLP 2020*.
- [10] Asai, A., et al. (2024). “Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection.” *ICLR 2024*.
- [11] Yan, S. Q., Gu, J. C., Zhu, Y., & Ling, Z. H. (2024). “Corrective Retrieval Augmented Generation.” *arXiv: 2401.15884*.
- [12] Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). “RAGAS: Automated Evaluation of Retrieval Augmented Generation.” *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 150-158.