

AnswerLink: 검색증강생성(RAG)을 활용한 최신 정보 기반 자연어 질의응답시스템

박성민, 이예빈, 민덕기*

건국대학교 컴퓨터공학부

holy.people.kr@gmail.com, dldpqls229@gmail.com, *dkmin@konkuk.ac.kr

AnswerLink: The recently information based Question Answering System using Retrieval Augmented Generation(RAG)

Seongmin Park, Yebeen Lee, Dugki Min*

Department of Computer Science and Engineering, Konkuk University

요약

최근 거대 언어 모델의 등장으로 이를 활용한 질의응답 시스템이 활발히 개발되고 있으나, 이미 학습된 정보의 갱신이 어렵고 환각 현상이 발생할 수 있다는 한계를 가지고 있다. 본 논문에서는 이러한 문제를 해결하기 위해, Retrieval Augmented Generation(RAG)을 기반으로 하여, 실시간으로 업데이트되는 뉴스 웹사이트의 크롤링을 통해 최신 정보를 반영하는 새로운 질의응답 시스템 AnswerLink를 제안한다. 또한, 형태소 분석을 통한 질의 및 문서 전처리 과정을 도입하여 검색기의 성능을 향상시켰다. 실험 결과, AnswerLink 시스템이 기존 모델과 비교하여 문서 검색 정확도 면에서 향상된 성능을 보였으며, 이는 중요 형태소만을 추출하는 전처리 과정이 크게 기여한 것으로 해석할 수 있다. 결론적으로 외부 지식 베이스의 구축과 형태소 분석 전처리를 통한 검색기의 성능 향상이 RAG구조를 이용하는 질의응답 시스템의 전반적인 성능 향상으로 이어질 수 있음을 시사한다.

I. 서론

질의응답 시스템은 사용자의 질문에 대한 답변을 제공하는 시스템으로, 다양한 분야에서 활용되고 있다. 특히, 최근 ChatGPT와 같은 거대 언어 모델의 등장으로 이를 활용한 오픈도메인 질의응답 시스템의 개발이 활발히 이루어지고 있다. 오픈도메인 질의응답 시스템은 질의에 대한 답이 포함된 문서를 사용자로부터 함께 제공받지 않고, 질의만을 입력받았을 때 적절한 응답을 제공해주는 시스템이다. 거대 언어 모델(Large Language Model, LLM)은 다양한 주제의 대용량의 문서집합을 학습해 파라미터로서 지식을 저장하고 있는 사전 학습 모델로, 외부 지식에 접근하지 않고 질문에 대한 답을 제공할 수 있다. 하지만 이미 학습된 지식에 대한 갱신이나 확장이 어려우며 환각 현상을 일으킬 수 있다는 한계점이 있다 [1]. 최근에는 언어 모델(Language Model, LM)의 크기는 더 이상 증가시키지 않으면서 새로운 정보와 지식을 쉽게 업데이트하기 위해 RAG(Retrieval Augmented Generation) 기법이 널리 활용되고 있다. RAG는 사용자 질의와 관련도가 높은 문서들을 선별하는 검색기(Retriever)와 검색기가 추출한 관련 문서들을 참고해 사용자 질의에 대한 응답을 생성하는 생성기(Generator)로 구성된 질의 응답 시스템 구조이다 [2].

본 논문에서는 기존 LLM 기반 질의 응답 시스템의 한계점을 극복하기 위해 RAG 기반의 시스템을 구축하고, 뉴스 웹사이트를 주기적으로 크롤링해와 정답을 추론하는데에 최신 정보 데이터를 외부 지식 베이스로 제공하는 새로운 시스템 구조(AnswerLink)를 제안한다. 이와 함께 형태소 분석기를 이용한 질의 및 문서 전처리를 통해 검색기의 성능 향상에 대한 연구를 진행하였다.

본 논문의 연구 과정은 다음과 같다. 제 II장에서는 본 연구에서 제안하는 AnswerLink 시스템의 전반적인 구조와 설계 세부사항을 설명하고 실험 및 성능 평가 결과를 제시한다. 제 III장에서는 본 연구의 의의 및 실험 결과를 종합하고 한계점 및 향후 연구 방향을 제안한다.

II. 본론

2.1 AnswerLink 시스템 설명

본 논문이 제안하는 AnswerLink 시스템은 크게 웹 서비스와 AI 모듈로 구성된다.

웹 서비스는 사용자와의 인터페이스를 담당한다. 웹 상단에는 최신 뉴스 헤드라인을 제시하고, 하단에는 사용자가 질문을 입력할 수 있는 칸과 전송버튼을 제공한다. 질문을 전송하면 채팅 형식으로 화면에 질문과 답변을 띄운다.

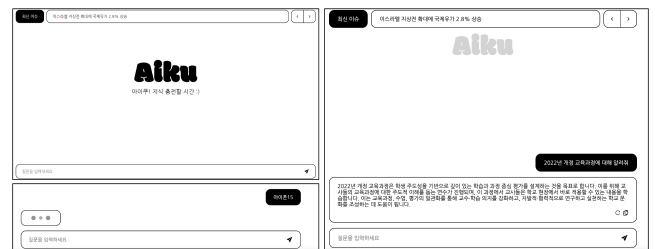


그림 1. 시스템 UI

AI 모듈은 사용자 질의에 대한 답변을 생성하는 역할을 한다. AI 모듈은 크롤러, 검색기, 생성기로 구성된다. 크롤러는 BIGKinds 주간이슈 웹사이트로부터 채팅창 상단에 제시할 최근 트렌드 뉴스 헤드라인을 읽어오며, Naver News 속보 웹사이트를 크롤링해 사용자 질의의 정답을 찾을 검색 대상 문서들의 집합인 외부 지식 베이스를 구축한다. 검색기로는 키워드 매칭 기반의 BM25 검색기를 사용하였다. 검색기는 전처리된 질의를 입력받아 관련도가 높은 상위 K개의 문서를 외부 지식 베이스로부터 선별한다.

이 때 검색기의 입력으로는 형태소 분석을 통해 사용자 질의에서 체언, 용언과 같이 중요한 형태소들('SL'(외국어), 'SH'(한자), 'SN'(숫자), 'NNP'(고유명사), 'NNG'(일반명사), 'NNB'(의존명사), 'NP'(대명사), 'NR'(수사), 'VV'(동사), 'VA'(형용사), 'VX'(보조용언), 'VCP'(긍정지정사), 'VCN'(부정지정사), 'XR'(어근))만 남긴 질의를 사용한다. 검색 대

상이 되는 외부 지식 베이스 문서들에 대해서도 형태소 분석을 수행하여 중요 형태소들만을 남긴 후 검색을 수행한다. 생성기로는 GPT를 사용하였다. 생성기는 입력 질의와 함께 검색기의 결과로 선별된 관련 최신 정보 문서를 전달받아 사용자 질의에 대한 답변을 생성한다. 전체 시스템에 대한 구성은 그림 2에서 도식화하였다.

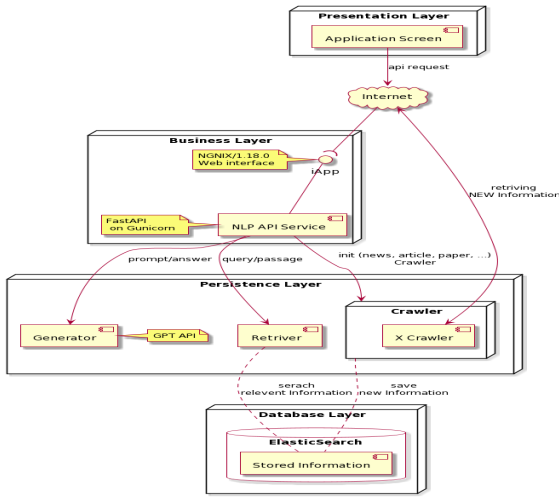


그림 2. AnswerLink 시스템 전체 구조

2.2 AnswerLink 시스템 성능 평가

AnswerLink 시스템의 생성기로 사용한 GPT는 기계독해를 잘 수행하지만 아무것도 주어지지 않은 상태(Zero-shot)에서는 환각 현상을 일으킬 수 있다는 문제가 있다 [1]. 이를 해결하기 위해 본 연구에서는 검색기의 정보 검색을 통한 정보 주입(Few-shot; In Context Learning) 방법을 이용하여 질의 검색 시스템의 전반적인 성능을 향상시키고자 하였다. 따라서 본 연구에서는 제안한 질의 응답 시스템의 성능 평가를 위해 검색기의 검색 정확도인 재현율(Recall)을 측정하였다 [3]. 평가 데이터 세트로는 KorQuAD v1.0의 학습 및 개발 데이터인 1,560개의 위키피디아 문서에 대한 66,181개의 질의응답 쌍을 사용하였다 [4].

관련 선행연구로 S. Cho et al.은 BM25 기반의 문서 검색기, 단락 선별기(LSTM, Transformer), 문서 독해기(BERT)로 구성된 질의 응답 시스템을 제안하였다 [5]. 선행연구와 본 연구는 동일하게 BM25 검색기를 사용했으나 검색기의 입력으로 선행연구는 원본 질의를 그대로 사용하는 반면, 본 연구에서는 형태소 분석을 통해 전처리된 질의를 사용했다는 점에서 차이가 있다. 검색하는 대상 문서에 대해서도 선행연구는 위키피디아 원본 문서를 그대로 사용하는 반면, 본 연구에서는 중요한 형태소들만 남긴 전처리된 문서를 사용했다는 차이점이 있다.

표 1에서는 검색한 상위 문서 개수인 Top K값에 따라 S. Cho et al.가 제안한 문서 검색기와 본 연구에서 제안한 AnswerLink 검색기의 문서 검색 정확도를 비교 제시하였다. 표 1의 결과를 그림 3에서 시각화하였다.

표 1. KorQuAD 데이터세트에 측정된 문서 검색 정확도(%)

Top K	5	10	20	50	100	200	300
선행연구	55.1	63.4	68.5	76.8	83.2	88.5	91.3
AnswerLink	96.3	97.8	98.8	99.4	99.6	99.8	99.8

표 1과 그림 3을 통해, 본 연구에서 제안한 방법을 적용하였을 때의 검색 정확도가 크게 향상된 것을 알 수 있다. 이는 검색기의 입력으로 전처리된 질의를 사용하였을 때 조사나 용언의 활용표현과 같은 불필요한 정보를

제거하여 질의의 의미를 보다 정확하게 파악할 수 있었기 때문이다. 조사는 주어, 목적어 등을 연결하는 역할을 하며, 용언의 활용 표현은 동사, 형용사의 의미를 구체화 하는 역할을 한다.

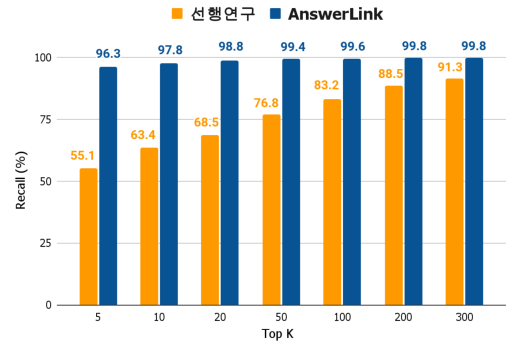


그림 3. 선행연구와 AnswerLink의 문서 검색 정확도 비교

이러한 조사나 용언의 활용 표현은 질의의 의미를 이해하는 데 중요한 역할을 하지만, 검색기의 입력으로 사용될 때는 오히려 혼란을 야기할 수 있다. 예를 들어, “2023년 12월 27일의 뉴스 헤드라인은 무엇인가요?”라는 질문을 검색할 때, 전처리된 질의는 “2023 12 27 뉴스 헤드라인”이 된다. 이때, 전처리된 질의는 불용어를 제거하여 질의의 의미를 보다 명확하게 파악할 수 있다. 특히 Top K가 5일 때의 검색 정확도 차이가 크게 나타나는데, 이는 본 연구에서 제안한 검색기가 관련있는 문서를 상위 순위에 배치하는데 효과적이었기 때문으로 보인다. 검색 결과의 개수가 적을수록 검색 정확도가 검색 결과의 순위에 더욱 민감하기 때문에, 관련있는 문서를 상위 순위에 배치하는 것은 검색 정확도를 높이는 데 중요한 요소이다. 검색 정확도의 향상은 질의 응답 시스템의 전반적인 성능 향상으로 이어진다. 본 연구가 제안하는 질의 응답 시스템은 검색된 문서를 바탕으로 답변을 생성하기 때문에, 검색 정확도가 높을 수록 보다 정확한 답변을 생성할 수 있다.

III. 결론

본 논문에서 제안한 AnswerLink 시스템은 RAG 기반 자연어 질의응답 시스템으로, 형태소 분석을 활용한 효과적인 전처리와 새로운 크롤링 메커니즘을 통해 기존 LLM의 한계를 극복하고자 하였다. 본 연구의 AnswerLink 검색기는 KoQuAD 데이터세트를 이용한 실험에서 높은 검색 정확도를 달성하였다. 이는 크롤링으로 구축한 외부 지식 베이스에 대한 검색을 통한 정보 주입 방법과, 중요 형태소들만을 추출함으로써 불필요한 정보를 제거하고 질의의 명확한 파악을 가능하게 한 전처리 때문이라고 해석할 수 있다.

향후 연구에서는 다양한 데이터 세트에 대한 성능 평가를 진행하여 시스템의 일반화 능력을 확인하고 다양한 도메인에서의 적용 가능성을 탐색할 것이다. 또한 BM25와 같은 언어적 일치를 기반으로 하는 검색기의 한계를 극복하기 위하여 LLM이 가지고 있는 기존 언어지식을 활용하여 질의를 확장하거나 재작성하는 질의 증강 과정을 통하여 검색기의 성능을 향상시키고자 한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A6A1A03046811, 2021R1A2C209494311). 또한 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술포럼원의 지원을 받아 수행된 연구임(P0020536, 2023년 산업혁신인재성장지원사업)

참 고 문 헌

- [1] OpenAI, "GPT-4 technical report. CoRR," 2023, abs/2303.08774.
- [2] Patrick L, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.
- [3] Ting, Kai Ming, "Encyclopedia of Machine Learning," Springer US, pp. 781-781, 2010.
- [4] S. Lim, M. Kim, J. Lee, "KorQuAD1.0: Korean QA Dataset for Machine Reading Comprehension," Journal of KIISE, pp. 539-541, 2018.
- [5] S. Cho, M. Kim, H. Kwon, "Korean Open Domain Question Answering System Using KorQuAD," Human and Language Technology, pp. 321-325, 2019.