

규칙 기반 AI 챗봇을 활용한 역사 교육에 관한 연구

김동주¹, 김나현¹, 김다희¹, 김현주¹, 고석주²

¹ 경북대학교 IT대학 컴퓨터 학부 학부생

² 경북대학교 IT대학 컴퓨터학부 교수

rlaehd62@knu.ac.kr, knh6110@knu.ac.kr, heesdajoy@knu.ac.kr,
rfmal15@knu.ac.kr, sjkoh@knu.ac.kr

A Study on History Education with Rule-Based Artificial Intelligence Chatbot.

Dong-Ju Kim¹, Na-Hyun Kim¹, Da-Hee Kim¹, Hyeon-Ju Kim¹, Seok-Joo Koh²

¹ Undergraduate Student, Department of Computer Science and Engineering, Kyungpook National University

² Professor, Department of Computer Science and Engineering, Kyungpook National University

요 약

본 논문은 현재 대구 경북지역을 대상으로 진행되고 있는 역사 교육에 있어 초·중등 학생들의 학습 활동을 진행하고 사용자의 질문에 정확한 질의응답을 제공하여 유용하게 활용될 수 있는 규칙 기반 AI 챗봇의 아키텍처와 이를 구성하는 각 단계를 제시한다.

1. 서론

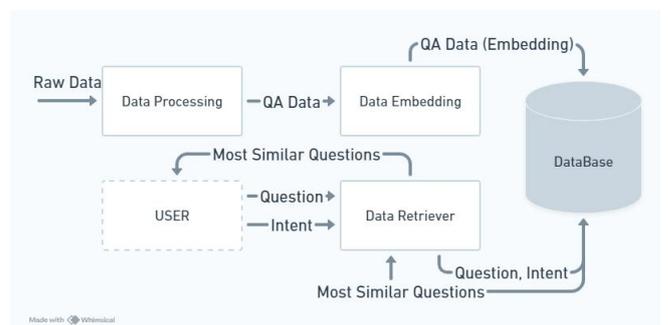
현재 대구·경북지역의 교육청은 학생들을 대상으로 지역사 교육을 진행하고 있다[1]. 그러나 유적지와 관련하여 수정된 정보가 반영되지 않은 교재를 활용하여 학습 활동을 진행하거나 유적지와 관련된 신뢰성 있는 정보를 제공받지 못하는 등의 다양한 문제들은 지역사 교육을 어렵게 만들고 있다. 이에 따라서 본 논문은 학생들의 원활한 학습 활동을 보조할 학습 도구의 개발을 위하여 규칙 기반의 인공지능 챗봇 아키텍처를 제시하고자 한다. 또한 본 논문의 프로젝트는 ㈜모티브와 함께 진행하는 경북대학교 산학협력 프로젝트의 일환으로, 실제 유적지가 구현된 메타버스와 연동될 예정임을 염두에 두고 구성됐다.

2. 본론

2.1 시스템 구조

본 논문은 역사 교육에 유용하게 활용될 수 있는 규칙 기반의 챗봇의 아키텍처와 이를 구성하는 핵심적인 처리 단계를 제시하는 것에 집중한다. 인공지능 챗봇의 아키텍처는 그림 1과 같이 3단계의 구성으로 이루어져 있다. 역사 자료를 기반으로 질의응답 데이터를 구축한 뒤 이를 임베딩 과정을 통해서

벡터 데이터로 변환한다. 이후 입력된 질문을 벡터 데이터로 변환한 다음, 입력된 의도로 분류된 질문들 중 가장 유사한 질문의 응답을 반환한다. 일반적인 규칙 기반의 인공지능 챗봇은 사용자의 질문과 가장 유사한 응답을 검색하기 위하여 의도 분류기를 사용한다. 그러나 본 논문은 해당 개념을 소개하고 설명하는 것에 목적을 두고 있지 않기 때문에, 사용자가 입력한 질문의 의도가 이미 분류되어 입력된다고 가정한다.



(그림 1) 시스템 아키텍처

2.2 시스템 처리 과정

2.2.1 데이터 구축

이 단계는 챗봇에서 사용될 질의응답 데이터를 구축한 뒤 이를 Intent를 중심으로 세부적으로 분류

한다. 본 논문은 산학협력 프로젝트의 일환으로, 멘토인 (주)모티브 측의 대구 역사 연구교사회의 자문을 통하여 연구 진행 및 질의응답에 필요한 역사 데이터를 확보하였다. 이후 역사 데이터를 기반으로 다음의 표 1과 같은 형태의 질의응답 데이터를 구축하였다.

2.2.2 데이터 임베딩

본 논문에서는 사용자의 질문과 가장 유사한 질

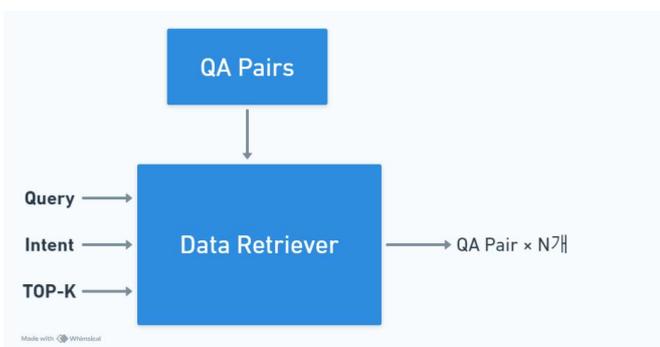
Intent	Question	Answer
대구읍성	대구읍성 메타버스에서 는 뭘 할 수 있어?	(답변 생략)
대구읍성	대구읍성이 뭐야?	''
석인상	석인상에 대해 자세히 알려줘	''
달성토성	달성토성이 왜 달성공 원이 되었어?	''

<표 1> 데이터 구성

의응답을 검색하기 위해, 구성된 질의응답 데이터를 Sentence Bert[2] 모델 ko-sroberta-multitask[3] 모델을 활용하여 임베딩 벡터 데이터로 변환하는 임베딩 과정을 거쳤다.

2.2.3 데이터 검색 및 반환

해당 단계는 임베딩 벡터 데이터가 추가된 질의응답 데이터를 기반으로 진행된다. 사용자가 입력한 의도(Intent)로 분류되는 질문들의 임베딩 데이터와 명시된 Sentence Bert 모델을 사용하여 사용자가 입력한 질문을 임베딩 벡터 데이터로 변환한다. 다음으로, 각각의 준비된 질문에 대해 코사인 유사도를 계산하여 입력된 질문과 가장 유사한 상위 N개의 질의응답 쌍을 사용자에게 반환한다.



(그림 2) 데이터 검색 및 반환

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

(그림 3) 코사인 유사도

3. 결론

본 논문에서는 역사 교육에 활용될 수 있는 위와 같은 구조의 규칙 기반 AI 챗봇의 아키텍처를 제시하였다. 현재 대중적으로 활용되고 있는 LLM의 고질적인 문제인 할루시네이션[4]은 금융, 법률, 교육 등의 분야에서는 치명적인 단점으로 작용한다. 질문의 의도를 분류하고 해당 의도를 중심으로 질문을 세부적으로 분류한 뒤 그중 가장 유사한 질문을 찾아내는 방식의 규칙 기반의 AI 챗봇을 활용한다면, 이를 통해 정확한 내용을 전달하여 학습과 관련 활동을 적절히 보조하여 효과적인 교육을 진행할 수 있다. 이러한 방식의 AI 챗봇은 특정 도메인을 기반으로 제공되는 챗봇 서비스에서 크게 효과적이며, 기존의 형식적인 교육에 비해 학습자(학생)들의 적극적인 참여를 이끌어내어 관심과 흥미를 유발하는 등, 본 논문에서 언급된 지역사에 관련된 학습 활동의 진행에 있어 크게 도움이 될 것으로 기대한다.

ACKNOWLEDGEMENT

“이 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 통해서 수행되었음 (2021-0-01082)“

참고문헌

[1] <https://www.daegubook.com/>
 [2] Reimers, Nils and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.” ArXiv abs/1908.10084 (2019)
 [3] github.com/jhgan00/ko-sentence-transformers
 [4] <https://it.donga.com/105162/>