

<https://doi.org/10.7236/JIIBC.2024.24.2.87>  
JIIBC 2024-2-13

## 검색 증강 LLM을 이용한 치과 상담용 챗봇 개발

# Development of Dental Consultation Chatbot using Retrieval Augmented LLM

박종진\*

Jongjin Park\*

**요약** 본 논문에서는 치과 상담용 챗봇을 개발하기 위해 기존의 대규모 언어 모델(LLM)과 랭체인 라이브러리를 이용하여 검색 증강 생성(RAG) 시스템을 구현하였다. 이를 위해 국내 치과 대학병원의 웹페이지 게시판에 있는 내용을 수집하고 치과 전문의의 자문과 감수를 받아 상담 데이터를 구축하였다. 입력된 상담용 데이터를 적절한 크기로 분할하기 위해 청크 크기와 각 청크의 겹치는 텍스트의 크기는 각각 1001과 100으로 하였다. 시뮬레이션 결과 검색 증강 LLM은 사용자 입력과 가장 유사한 상담 내용을 검색하여 출력하였다. 구축된 챗봇을 통해 치과 상담의 접근성과 상담 내용의 정확성을 높일 수 있음이 확인되었다.

**Abstract** In this paper, a RAG system was implemented using an existing Large Language Model (LLM) and Langchain library to develop a dental consultation chatbot. For this purpose, we collected contents from the webpage bulletin boards of domestic dental university hospitals and constructed consultation data with the advice and supervision of dental specialists. In order to divide the input consultation data into appropriate sizes, the chunk size and the size of the overlapping text in each chunk were set to 1001 and 100, respectively. As a result of the simulation, the Retrieval Augmented LLM searched for and output the consultation content that was most similar to the user input. It was confirmed that the accessibility of dental consultation and the accuracy of consultation content could be improved through the built chatbot.

**Key Words** : Chatbot, Dental consultation, LLM, RAG

### 1. 서론

최근 대규모언어모델(Large Language Model, LLM)이 화두가 되고 있다. 2018년 OpenAI에서 GPT-1 (Generative Pretraining Transformer)가 발표된 이

후 2019년 GPT-2, 2020년 GPT-3에 이어 다양한 LLM이 발표되고 있으며 2022년 GPT-3.5와 여기에 RLHF (Reinforcement Learning From Human Feedback)를 적용한 ChatGPT가 공개되었고, 2023년에는 GPT-4가 발표되었다. 이후 LLM은 최신 소프트웨어 스

\*정회원, 청운대학교 컴퓨터공학과  
접수일자 2024년 3월 5일, 수정완료 2024년 3월 30일  
게재확정일자 2024년 4월 5일

Received: 5 March, 2024 / Revised: 30 March, 2024 /  
Accepted: 5 April, 2024

\*Corresponding Author: jjpark@chungwoon.ac.kr  
Dept. of Computer Engineering, Chungwoon University,  
Incheon, Korea

택의 한 요소로 빠르게 부상하고 있으며 OpenAI, 메타, 구글과 같은 빅테크 기업이 제공하는 모델 API나 개인화된 맞춤 모델(custom model)을 이용한 LLM 기반 애플리케이션을 구축하려는 다양한 시도가 시도되고 있다. 최근에는 GPT-4V와 같은 다양한 형태의 데이터 형태로 학습한 대규모 멀티모달모델(Large Multi-Modal Model, LMM)로 진화하고 있다<sup>[1]</sup>.

LLM은 대규모 데이터 셋으로 훈련된 인공지능 언어 모델로서 원래 자연어처리(Natural Language Processing, NLP)에 사용되기 위해 개발되었고 텍스트 생성, 분류, 번역, 질의응답 및 감정 분석 등 다양한 작업을 수행할 수 있다. 일반적으로 LLM은 수천만 개 이상의 가중치 매개변수를 가지고 있으며 매개변수의 개수가 많아질수록 창발(Emergency) 능력이 나타나는 것이 증명되면서 매개변수의 수가 급증하고 있다. GPT-3.5의 경우 1750억 개의 매개변수를 가진다고 알려져 있고, 공개되지 않았지만 GPT-4는 수조 개의 매개변수를 가진 것으로 추정된다.

각 기업이나 개인들은 이러한 LLM을 자신들의 응용 프로그램에 통합하는 것을 고려하기 시작했고 다양한 시도들이 진행되어 왔다. 그러나 단순히 LLM API를 중심으로 래퍼(wrapper)를 개발하는 경우 추가적인 가치를 제공하지 않기 때문에 애플리케이션이 성공하지 못할 가능성이 높다. LLM의 주요 문제 중 하나는 지식 단절(knowledge cutoff)이다. 예를 들어 웹 기반 챗봇 형태의 ChatGPT에 2023년에 일어난 것에 대해 문의하면 제대로 응답하지 못한다.

또한 LLM을 학습할 때 사용한 학습 데이터셋에 없는 이벤트에 대해 LLM에 질문하는 경우에도 동일한 문제가 발생한다. LLM은 지식 단절 일(cutoff date) 이후에는 공개적으로 이용 가능한 모든 정보에 대해 지식이 없지만, 지식 단절 일 이전에도 이용 가능할 수 있는 개인 정보 또는 기밀 정보에 대한 지식은 얻을 수 없다. 예를 들어, 특정 회사의 기밀 정보는 일반 데이터를 학습한 LLM은 접근할 수 없으므로, 이런 경우 맞춤형 LLM을 통한 질문-응답(Q&A) 시스템 개발이 필요하다<sup>[2]</sup>.

LLM의 또 다른 문제는 정확하지 않을 수 있는 사실적인 텍스트를 생성하도록 훈련되었다는 것으로 이를 환각 현상(Hallucination)이라고 한다. 이 때 일부 유효하지 않은 정보는 다른 정보보다 발견하기가 더 어렵다. 특히 누락된 데이터의 경우, LLM이 학습에서 기본 사실이 부족하다는 것을 인정하는 대신 설득력 있게 들리지만 그럼에도 불구하고 잘못된 답변을 구성할 가능성이 매우

높다. 따라서 LLM의 지식을 업데이트하고 확장하는 것은 오늘날 매우 적절한 과제 중에 하나이다.

오늘날 LLM의 영향을 생각할 때, 이와 같이 LLM의 답변을 확인하거나 보다 정확한 결과를 생성하는 것은 해결해야 할 큰 문제이다. 이 문제의 해결 방법은 크게 두 가지가 있다. 하나는 미세조정(fine-tuning)이고, 다음은 검색 증강 LLM(retrieval-augmented LLM)을 구축하는 것이다. LLM을 미세조정 하는 경우, 비용이 많이 들고 새로운 훈련 데이터셋을 구축해야 하는 어려움이 있다. 또한, 실시간 응답이나 개인적인 정보를 포함하는 경우 제대로 응답하지 못한다<sup>[3]</sup>. 검색 증강 생성(Retrieval Augmented Generation, RAG) 방식은 LLM을 사용하여 데이터 소스에서 추가로 제공된 관련 문서를 기반으로 답변을 생성하는 방식으로 이러한 문제들을 해결할 수 있다.

의료 영역에서 질병 치료와 과정에 대한 상담은 매우 중요한 부분으로 치과에서도 치료 전 상담의 필요가 증가하고 있다. 치과 상담은 많은 사람들에게 중요한 건강 관리 부분 중 하나로 자리잡았다. 치아와 구강 건강은 개인의 일반적인 건강과 삶의 질에 큰 영향을 미치며, 치과 질환은 미룰 수 없는 문제가 될 수 있다. 그러나 치료 및 상담에 대한 접근성은 전 세계적으로 일관되게 낮고, 이로 인해 많은 사람들이 필요한 치료 관리를 받지 못하고 있다. 한편, 정보 기술 분야의 혁신은 의료 분야에도 큰 영향을 미치고 있다. 특히, 인공지능(AI)과 자연어처리 기술의 발전은 새로운 의료 서비스 및 의료 관리 방법을 제공하고 있으며, 치과 상담 역시 이러한 혁신의 이점을 누릴 수 있는 분야 중 하나이다. 대규모언어모델(LLM)을 이용한 치과 상담용 챗봇은 환자들에게 효과적인 정보와 조언을 제공하며, 치과 전문가들과의 의사소통을 보조하는 강력한 도구가 될 수 있다<sup>[4]</sup>.

본 논문은 오픈AI에서 발표한 LLM인 GPT3.5를 파운데이션 모델로 사용하여 치과상담용 Q&A 챗봇을 개발하기 위해 관련 치과 상담 데이터를 구축하고 이를 이용하여 검색 증강 생성(RAG) 시스템을 구축하였다. 구축된 RAG LLM을 통해 치과 상담의 접근성과 상담 내용의 정확성을 높이고 치과 질환 예방 및 관리에 도움을 주는 가능성을 탐구한다. 치과 상담의 필요성과 이를 충족시키기 위해 최근 급속하게 발전하고 있는 LLM과 검색 증강 생성 기술의 활용에 대한 논의를 통해, 본 논문은 치과 건강을 개선하고 개인의 삶의 질을 향상시키는 연구 주제를 제시하고자 한다.

## II. 검색 증강 생성 LLM

### 1. 대규모 언어 모델(LLM)

ChatGPT와 같은 LLM은 인간과 같이 텍스트를 이해하고 생성할 수 있는 능력으로 인해 2022년 발표된 이후로 세계를 놀라게 하고 있다. 즉, 다양한 대화 맥락에 적응하고, 광범위한 주제에 대한 질문에 답하고, 창의적인 글쓰기를 시뮬레이션하는 능력은 인간과 기계가 상호 작용하는 방식에 혁명을 일으켜 인공지능 애플리케이션의 새로운 물결을 촉발시켰다<sup>[5]</sup>.

이러한 혁명에도 불구하고 LLM을 이용하여 다양한 분야에 응용할 때 한계도 존재한다. 그것은 첫째, 지식 단절(knowledge cutoff), 둘째, 학습데이터의 제한, 그리고 셋째, 환각 현상(Hallucination) 등이다. 지식 단절이란 LLM이 학습 데이터셋(training dataset)을 이용한 학습 후에 발생한 어떤 이벤트에 대해서도 답변하지 못한다는 것을 의미한다. GPT-3.5와 GPT-4는 2021년 9월까지의 데이터로, GPT-4 turbo는 2023년 4월까지의 데이터로 각각 학습하여 이 시점 이후의 질문에 대해서는 제대로 답변하지 못하고 잘못된 답변을 생성할 가능성이 있다. 학습데이터의 제한에 대해서는 LLM 학습 시 개인 정보 또는 기밀 정보에 대한 지식은 얻을 수 없으므로 특정 기업을 위한 Q&A 챗봇 구축과 같은 경우에는 맞춤형 LLM(custom LLM)의 구축을 통해 해결할 수 있다. 환각 현상은 LLM이 답변을 생성 시 지식의 한계를 인정하는 대신 사실과 다른 그럴듯한 잘못된 답변을 생성하는 것이다.

이러한 문제들의 해결 방법은 크게 두 가지가 있다. 하나는 미세조정(fine-tuning)이고, 다음은 검색 증강 LLM(Retrieval-Augmented LLM)을 구축하는 것이다. 미세조정은 사용자만의 데이터를 이용하여 LLM을 자신의 용도에 맞게 맞춤화하는 것으로 이를 통해 모델의 성능을 향상시킬 수 있으며, 명령어를 더 잘 따르게 하거나, 특정 언어로 항상 응답하도록 만들 수 있다. 또한, 미세조정을 통해 모델의 출력 형식을 일관되게 만들거나, 특정 브랜드의 목소리와 일치하게 할 수 있다. 자신들이 가진 데이터에 대한 유출 문제도 해결될 수 있다<sup>[2]</sup>. 그러나 이 방법은 대규모의 파라미터를 가진 언어모델을 학습시키기 위해 새로운 학습 데이터셋을 구축해야 하고 GPU 등이 장착된 서버급 컴퓨터가 필요하며 이에 따라 비용이 많이 사용되는 어려움이 있다.

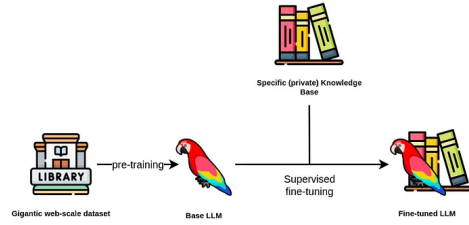


그림 1. 미세조정 개념도<sup>[2]</sup>  
 Fig. 1. Concept of fine-tuning

### 2. 검색 증강 생성 (Retrieval Augmented Generation)

일반적인 데이터로 학습된 LLM의 한계를 극복하는 또 다른 방법인 검색 증강 생성은 LLM과 정보검색시스템을 결합하는 방식이다. 이것은 언어모델이 사용자의 질문에 답변할 때 필요한 정보를 외부 데이터베이스에서 검색할 수 있도록 하며, 이를 통해 더욱 정확하고 필요한 답변을 제공할 수 있게 된다. 즉, 학습데이터에 포함되지 않은 또는 보안상 공개되지 않은 내부 지식 데이터에 대해 기존의 LLM을 사용하는 대신, 검색 증강 LLM을 사용하는 것이다. 여기서는 LLM을 회사나 개인 정보의 자연어 인터페이스로 사용한다<sup>[2]</sup>.

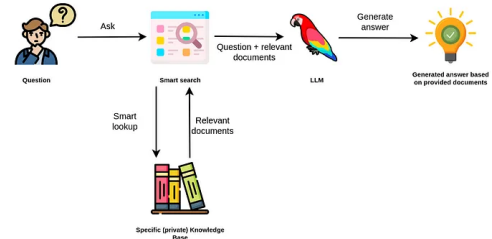


그림 2. 검색 증강 생성의 개념도<sup>[2]</sup>  
 Fig. 2. Concept of RAG

RAG을 구현하는 방식은 랭체인(Langchain)과 라마 인덱스(LlamaIndex)를 이용하는 방법이 있다. 랭체인은 대규모언어모델(LLM)을 기반으로 한 어플리케이션 개발을 위한 프레임워크로, 언어 모델을 사용한 다양한 작업을 효과적으로 수행할 수 있게 해주는 도구이다. 라마 인덱스는 랭체인 라이브러리를 기반으로 구축된 어플리케이션이다. 본 논문에서는 사용자 인터페이스가 더 간편하고 사용하기 쉬운 랭체인을 사용한다. RAG 어플리케이션은 주요한 두 가지 구성 요소로 이루어진다. 즉, 인덱싱과 검색 및 생성이다. 인덱싱은 소스에서 데이터를

수집하고 인덱싱하기 위한 파이프라인으로 일반적으로 오프라인에서 이루어진다. 검색 및 생성은 런타임 시 사용자 쿼리를 가져와 인덱스에서 관련 데이터를 검색한 다음 이를 언어모델에 전달하는 실제 RAG 체인이다<sup>[6]</sup>. 그림 3은 랭체인을 이용한 RAG의 구현 개념도를 나타낸다. 여기서 인덱싱은 로드(Load), 분할(Split) 그리고 저장(Store)으로 이루어져 있고 랭체인에서는 각각 Document Loaders, Text Splitters, Vector Embeddings를 이용하여 이를 구현한다. 검색과 생성은 각각 Retrievers와 LLM에 의해 구현된다. 사용자의 데이터가 인덱싱되어 수치화된 데이터가 벡터 저장소에 저장된 후, 사용자의 질문이 입력되면 사용자의 질문과 가장 유사한 부분(Chunk)이 벡터 저장소에서 검색기(Retrievers)에 의해 검색되고 언어모델(LLM)에 넘겨진다. LLM은 사용자의 질문과 검색된 데이터를 포함한 프롬프트(Prompt)를 이용하여 적절한 답변을 생성하여 출력한다.

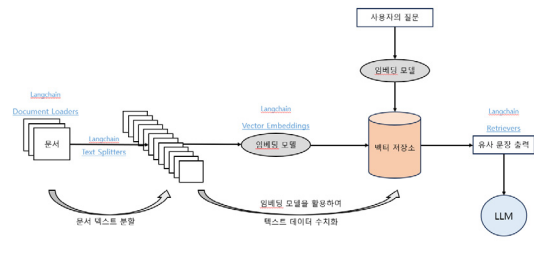


그림 3. 랭체인을 이용한 RAG의 구현 개념도  
Fig. 3. Concept of RAG implementation using Langchain

### III. 치과 상담용 데이터셋

랭체인을 이용한 검색 증강 LLM을 구축하고 이를 이용하여 치과 상담 챗봇을 구현하기 위해 치과 상담 데이터셋을 만들어야 한다. 이를 위해 국내 치과대학병원의 웹페이지 게시판에 있는 내용을 수집하여 상담에 부적절한 내용을 제거하고 적절한 크기의 데이터셋으로 수정 및 편집하는 과정을 거쳤다<sup>[7-9]</sup>. 이 과정에서 치과 전문의의 자문과 감수를 받아서 데이터를 최종적으로 확정하였다.

치과대학병원의 치과상식, 구강건강정보 또는 진료상담FAQ 게시판의 주제와 내용은 다양한 주제에 대해 진료과 별로 정확한 내용을 포함하고 있다. 그러나 때때로

치과상담과 관련성이 적은 내용을 포함하거나 내용이 너무 길어 검색 증강 LLM을 구축하기에 부적절한 경우 내용을 적절하게 수정하고 편집하였다. 그림 4는 수정·편집 전의 게시판 내용과 수정·편집 후의 데이터셋을 보여준다. 그림에서 노란색으로 표시한 원데이터의 텍스트는 치과대학병원 게시판의 제목으로 상담의 주제이다. 한 주제의 내용이 다소 긴 경우는 그림과 같이 좀 더 정확하고 다양한 내용에 대한 상담을 위해 치과 전문의의 자문을 받아 좀 더 세분화된 주제의 데이터 덩어리(chunk)로 나누었다. 구축된 치과상담용 챗봇 데이터는 60218 글자로 이루어져 있다.

#### 교정 치료와 치아 관리

교정 치료를 시작하게 되면 치아 표면에 브라켓이라고 하는 고정 장치를 부착하게 되는데, 이 브라켓을 정형 철사가 삽입되는 슬롯이라고 하는 홈과 브라켓 링이라고 하는 작은 갈고리같이 생긴 구조물로 이루어진다. 브라켓을 부착할 때에는 고정용 레진이라고 하는 재료를 사용하는데 치아 표면에 레진을 올리고 브라켓이 치아에 부착되게 된다. 고정치료가 끝나게 되면 브라켓을 치아에서 제거해야하므로 레진의 일부는 고정용 레진의 강도는 충치 치료를 위해 사용하는 레진의 강도보다는 낮다. 따라서, 딱딱한 음식이나 식을 무신고 씹게 되면 치아에 부착된 브라켓이 탈락할 수 있다. 어느 한 치아에서만 브라켓이 탈락하였던 방향으로 진행될 수 있으므로 브라켓이 떨어진 치과에 바로 내원하여 다시 부착해야 한다. 간혹 치과에 내원하는 환자들이 있는데 그렇게 되면 한 달 동안 나머지 치아들만 이동을 하고, 브라켓이 떨어진 치아는 치과에 다시 천단계로 들어가야 해서 치료기간이 더 연장될 수 있다.

치아에 브라켓을 부착하고 고정치료가 시작되면 환자는 바로 불편감을 느끼게 된다. 예전과는 다르게 사시에 음식물이 자주 끼기 때문이다. 따라서 고정치료를 시작하게 되면 치아관리에 보다 더 주의를 기울여야 한다. 구강세정기는 작고 긴 노즐을 통해 물줄기를 입안으로 뿜어내어 철사나 브라켓 사이사이에 음식물 찌꺼기에 도움이 되는 효과가 있다. 수압의 강도는 조절할 수 있으며 처음부터 너무 세 강도보다는, 이처럼 구강세정기는 고정장치 주변에 끼여 있는 음식물 제거와 잇몸관리에는 도움이 되지만 궁극적으로는 정도의 육적용을 가지고 사용한 후 잇솔질을 해야 한다. 고정용 잇솔을 사용해도 되고 일반 칫솔은 가운데 부분이 오목하게 들어가 있어 잇솔질 할 때 고정장치가 닿을 수 있게 디자인되어 있지만; 예 큰 문제는 없다. 치약은 많이 사용하지 않으면 상쾌한 느낌을 가질 수 있으나 거품이 많이 나서 크기 정도가 적당하다.

교정치료를 받는 동안에는 식사 후 바로 양치질을 하는 습관을 가지는 것이 좋다. 고정장치 주변에 음식물 찌꺼기, 음식물 찌꺼기와 치약을 휴대하여 수시로 양치질을 할 수 있도록 하는 것이 좋고 경우에 따라 구강세정기도 필요하다. 이렇게 고정치료를 받는 1-2년 동안 양치질 습관이 갖추어지면 고정치료가 끝난 후에도 그 습관이 유지되면 구강건강에 한 달 정도는 계속적으로 사용하는 것이 좋다. 고정치료가 끝난 직후에는 일시적으로는 잇몸을 적절히 자극해주는 구강세정기를 사용해주면 도움이 될 수 있다. 무엇보다도 고정치료를 받는 동안, 정기적인 내원 약속을 잘 지켜야 치료가 원활히 이루어질 것이다. 간혹 개인 사정으로 못 갈지라도 교정치료의 부작용이 생길 수 있다. 긴 치료 기간 동안 치아관리에 주의를 기울이는 한편 정기적으로 검사와 가지런한 치아를 가지는 기쁨을 얻을 수 있을 것이다.

(a) 수정·편집 전 원데이터

#### 교정 치료 시 딱딱한 음식이나 끈적거리는 음식은 먹으면 안되는 이유

교정 치료를 시작하게 되면 치아 표면에 브라켓이라고 하는 고정 장치를 부착하게 되는데, 이 브라켓을 모든 치아에 부착하고 짧게는 1년 길게는 2년 넘게 지내야 한다. 브라켓은 고정용 철사가 삽입되는 슬롯이라고 하는 홈과 브라켓 링이라고 하는 작은 갈고리같이 생긴 구조물로 이루어져 있다. 치아 하나하나에 부착되어야 하므로 브라켓의 크기는 기본적으로 작다. 브라켓을 부착할 때에는 고정용 레진이라고 하는 재료를 사용하는데 치아 표면에 레진을 올리고 브라켓을 눌러 부착시킨 다음 LED 광원을 쬐어주면 ...

#### 교정 치료 시 구강위생은 어떻게 관리하나요?

치아에 브라켓을 부착하고 고정치료가 시작되면 환자는 바로 불편감을 느끼게 된다. 예전과는 다르게 입안에서 이물감이 느껴지고 브라켓과 철사가 모든 치아에 부착되어 있어 식사 시에 음식물이 자주 끼기 때문이다. 따라서 고정치료를 시작하게 되면 치아관리에 보다 더 주의를 기울여야 한다. 먼저 워터픽 또는 구강세정기라고 하는 구강용품을 사용하는 것을 추천한다. 구강세정기는 작고 긴 노즐을 통해 물줄기를 입안으로 뿜어내어 철사나 브라켓 사이사이에 끼여있는 음식물 찌꺼기를 제거해준다. 또한 수압으로 잇몸을 자극하여 잇몸의 혈액순환에 도움이 되는 효과가 있다. ....

#### 교정 치료 시 구강세정에 힘써야 하는 이유

교정치료를 받는 동안에는 식사 후 바로 양치질을 하는 습관을 가지는 것이 좋다. 고정장치 주변에 음식물이 끼어 있으면 보기에 좋지 않고 입 냄새의 원인이 될 수 있기 때문이다. 따라서 항상 칫솔과 치약을 휴대하여 수시로 양치질을 할 수 있도록 하는 것이 좋고. ....

(b) 수정·편집 후 치과 상담용 데이터

그림 4. 치과상담용 데이터의 수정 및 편집  
Fig. 4. Modification and editing of data for dental consultation

## IV. 시뮬레이션 및 결과

구축된 데이터셋을 로딩(loading)하고 분할(split)하기 위해 랭체인(LangChain)의 Document Loaders와 Text Splitters 라이브러리를 사용하였다. 분할된 청크(chunk) 데이터는 Vector Embeddings를 이용하여 벡터 저장소에 저장하였다. 데이터를 분할하기 위한 파라미터는 아래와 같다.

- chunk\_size=1001,
- chunk\_overlap=100

여기서, 청크 크기(chunk\_size)는 텍스트를 분할하는 글자 수의 최대 크기를 의미한다. 청크 오버랩(chunk\_overlap)은 청크 텍스트간의 겹치는 내용의 최대 크기(글자 수)를 의미한다. 이는 내용의 단절로 인해 부정확한 출력을 생성하는 것을 방지하기 위해 적절한 값을 사용한다.

구축된 데이터를 분할한 결과는 그림 5와 같다. 입력된 데이터가 적절하게 잘 분할된 것을 볼 수 있다. 그림 4 (b)의 첫 번째, 세 번째 단락의 데이터가 출력변수 docs[0]와 docs[2] 배열에 입력되어 있는 것을 볼 수 있다.

```
1 | 1 print(docs[0])
2 | 2 print("-" * 100)
3 | 3 print(docs[1])
4 | 4 print("-" * 100)
5 | 5 print(docs[2])

page_content="고정 치료 시 딱딱한 음식이나 끈적거리는 음식을 먹으면 안되는 이유?<math>K_{max}</math>값을 시작하게 되면
page_content="머글나는 안쪽이라 잘 안맞이는데 머글나의 충치 예방에는 어떻게 효과적일까요?<math>K_{max}</math>머글나는 안쪽에
page_content="고정 치료 시 구강 세정에 힘써야 하는 이유?<math>K_{max}</math>값을 받는 동안에는 식사 후 바로 말치질 하
```

그림 5. 분할된 결과  
 Fig. 5. Split Results

사용자가 질문을 입력하면 분할된 데이터를 이용하여 가장 유사도 높은 텍스트를 찾아 출력한다. 그림 6은 상담용 원데이터 청크를 나타내고, 그림 7은 상담 데이터에 있는 주제와 같은 질문(query)을 입력하고 질문과 가장 유사도가 가까운 텍스트 청크를 출력한 결과이다. 원데이터의 내용과 같은 내용을 출력하는 것을 볼 수 있다.

**유음 어린이들은 불소치료를 많이 받는데 불소가 충치 예방에 도움이 되나요?**  
 불소도포를 하면 치아가 단단해져 충치에 대한 저항성을 갖게 되고, 입안 세균의 증식을 억제하여 충치예방에 강한 아이보다 취약한 아이가 보약을 먹으면 더 큰 효과를 얻을 수 있는 것처럼, 충치가 없는 아이보다 구강과 수 있습니다. 불소 치료는 먼저 치아표면을 깨끗하게 한 다음, 탈염알칼리 용액이다 불소 연고를 담아서 아이에게 소연고를 발라주게 됩니다. 불소치료 후 한 시간동안은 물도 마시지 않도록 합니다. 불소도포의 시기는 충치가 없는 어린이는 6달 간격을 추천하고 있습니다.

그림 6. 치과 상담용 원데이터  
 Fig. 6. Raw data for dental consultation

```
7 # 질문 쿼리
8 query = "유음 어린이들은 불소치료를 많이 받는데 불소가 충치 예방에 도움이 되나요?"
9
10 query_embedding = model.encode(query) # 쿼리의 임베딩 계산
11
12 # sep 리스트의 각 텍스트 조각에 대한 임베딩 계산
13 sep_embeddings = [model.encode(content[i]) for content in sep]
14
15 # 쿼리 임베딩과 각 텍스트 조각 임베딩 사이의 코사인 유사도 계산
16 query_embedding = np.array([query_embedding])
17 cosine_scores = [pytorch_cos_sim(query_embedding, np.array([emb])) for emb in sep_embeddings]
18
19 # 가장 유사도가 높은 텍스트 조각의 인덱스 찾기
20 cosine_scores_np = np.array([score.item() for score in cosine_scores])
21
22 most_similar_index = np.argmax(cosine_scores_np)
23
24 # 가장 유사한 텍스트 조각 출력
25 print(sep[most_similar_index][0])
26
```

유음 어린이들은 불소치료를 많이 받는데 불소가 충치 예방에 도움이 되나요?  
 불소도포를 하면 치아가 단단해져 충치에 대한 저항성을 갖게 되고, 입안 세균의 증식을 억제하여 충치예방에

그림 7. 유사도평가에 따른 출력 결과  
 Fig. 7. results according to similarity evaluation

## V. 결론

본 논문에서는 치과상담용 Q&A 챗봇을 개발하기 위해 기존의 LLM과 랭체인 라이브러리를 이용하여 검색 증강 생성(RAG) 시스템을 구현하였다. 이를 위해 국내 치과대학병원의 웹페이지 게시판에 있는 내용을 수집하고 치과 전문의의 자문과 감수를 받아 Q&A 챗봇에 사용할 데이터를 구축하였다. RAG 애플리케이션의 두 가지 구성 요소인 인덱싱과 검색·생성은 랭체인에서 제공하는 로더, 분할기, 벡터 임베딩 그리고 검색기(Retrievers)를 사용하여 구현하였다. 입력된 상담용 데이터를 적절한 크기로 분할하기 위해 청크 크기와 겹치는 텍스트의 크기는 각각 1001과 100으로 하였다. 시뮬레이션 결과 사용자 입력과 가장 유사한 내용의 텍스트를 검색하여 출력하였다. 구축된 RAG LLM을 통해 치과 상담의 접근성과 상담 내용의 정확성을 높이고 치과 질환 예방 및 관리에 도움을 줄 수 있음이 확인되었다.

## References

- [1] <https://openai.com/>
- [2] Knowledge Graphs & LLMs: Fine-Tuning Vs. Retrieval-Augmented Generation, <https://medium.com/neo4j/knowledge-graphs-llms-fine-tuning-vs-retrieval-augmented-generation-30e875d63a35>
- [3] W. X. Zhao, et. al., "A Survey of Large Language Models," *arXiv preprint arXiv:2303.18223*, 2023. <https://doi.org/10.48550/arXiv.2303.18223>
- [4] An overview of oral health promotion in adolescents, "International Journal of Paediatric Dentistry, Vol. 19, No. 3, pp. 163-71, 2008.

<https://doi.org/10.1111/j.1365-263X.2008.00954.x>

- [5] Knowledge Graphs & LLMs: Harnessing Large Language Models with Neo4j,  
<https://medium.com/neo4j/harnessing-large-language-models-with-neo4j-306ccbdd2867>
- [6] [https://python.langchain.com/docs/use\\_cases/question\\_answering/](https://python.langchain.com/docs/use_cases/question_answering/)
- [7] <https://dent.khmc.or.kr/kr/health-knowledge-dental-common/list.do>
- [8] <https://www.cnudh.co.kr/board/B0051.cs?m=40>
- [9] [https://www.snudh.org/portal/bbs/selectBoardList.do?bbsId=BBSMSTR\\_000000000258&menuNo=25010000](https://www.snudh.org/portal/bbs/selectBoardList.do?bbsId=BBSMSTR_000000000258&menuNo=25010000)
- [10] Zhiheng Xi, et. al., "The Rise and Potential of Large Language Model Based Agents: A Survey," *arXiv preprint arXiv:2309.07864*, 2023.  
<https://doi.org/10.48550/arXiv.2309.07864>
- [11] S. Y. Hwang, et. al., "Big Data-based Medical Clinical Results Analysis," *JIIBC*, VOL. 19 NO. 1, Feb. 2019.  
DOI : 10.7236/JIIBC.2019.19.1.187
- [12] Youngjun Jo, Jongwon Kim, "A Study on The Classification of Target-objects with The Deep-learning Model in The Vision-images," Vol. 22, No. 2, pp. 20-25, 2021.  
<http://dx.doi.org/10.5762/KAIS.2021.22.2.20>
- [13] Dong-Hwan Shin, et. al., "Malware Detection Model-based on Natural Language Processing with Simplified Data Preprocessing," *Journal of KIIT*. Vol. 22, No. 2, pp. 1-8, Feb. 28, 2024.  
<http://dx.doi.org/10.14801/jkiit.2024.22.2.1>

## 저 자 소 개

### 박 종 진(정회원)



- 1989년 : 연세대학교 전기전자공학과 학사 졸업
- 1991년 : 연세대학교 전기전자공학과 석사 졸업
- 1997년 : 연세대학교 전기전자공학과 박사 졸업
- 현재 : 청운대학교 컴퓨터공학과 교수
- 관심분야 : 딥러닝, 영상처리, 임베디드시스템