

멀티모달 LLM기반 표/차트 데이터 직렬화를 이용한 검색/생성 능력 개선

손 현 규¹⁾·강 창 목²⁾·김 한 울^{*3)}

현대자동차¹⁾·한양대학교 전기공학과²⁾·서울과학기술대학교 인공지능응용학과³⁾

Enhanced Retrieval and Generation Performance via Multimodal LLM-based Serialization of Tabular and Chart Data

Hyun Kyu Shon¹⁾ · Changmook Kang²⁾ · Hanul Kim^{*3)}

Hyundai Motor Company¹⁾, Hanyang University²⁾, Seoul National University of Science and Technology³⁾

Key words : Multimodal LLM (멀티모달 대규모 언어모델), Visual Data Serialization (시각적 데이터 직렬화), Table & Chart Serialization (시각적 표 · 차트 직렬화), Retrieval-Augmented Generation (RAG) (검색 증강 생성), Document Preprocessing for Retrieval (검색을 위한 문서 전처리), Document Layout Analysis (문서 구조 분석)

교신저자, E-mail: hukim@seoultech.ac.kr

검색증강생성(RAG)은 대규모 코퍼스 기반 질의응답 시스템의 핵심 기술로서 주목 받고 있다. 그러나 표나 차트와 시각적·구조적 정보에 대해 RAG 파이프라인의 성능이 열화되어, 기업·공공기관과 같은 핵심 영역의 고가치 정보를 다루는데 기술적 제약이 발생한다. 기존의 광학문자인식(Optical Character Recognition, OCR)과 단순 텍스트 분할(chunking)에 기반한 접근은 데이터의 의미를 결정하는 행·열 관계 등 시각적 레이아웃에 내재된 구조적·의미적 맥락을 보존하지 못해 검색 정확도를 떨어트리고 환각(hallucination) 가능성을 높인다. 이에 본 연구는 멀티모달 대규모 언어모델(Multimodal Large Language Model, MLLM)을 활용하여 시각적 데이터를 지능적으로 직렬화(serialization)하는 2단계 파이프라인을 제안한다. 구체적으로 MLLM이 표와 차트의 시각적 구조를 해석한 후 서술적이고 의미적으로 풍부한 JSON 형식으로 변환함으로써 RAG에서의 검색 성능을 개선한다. 또한, 시각 언어 모델(Visual Language Model)이나 OCR 엔진 등 GPU 집약적 구성요소를 추가하지 않고, 단일 MLLM의 통합 영상 처리 능력을 활용해 기술 스택을 단순화하고 운영 복잡성을 낮추었다. 실험결과는 본 접근 방법의 우수성을 확인한다.

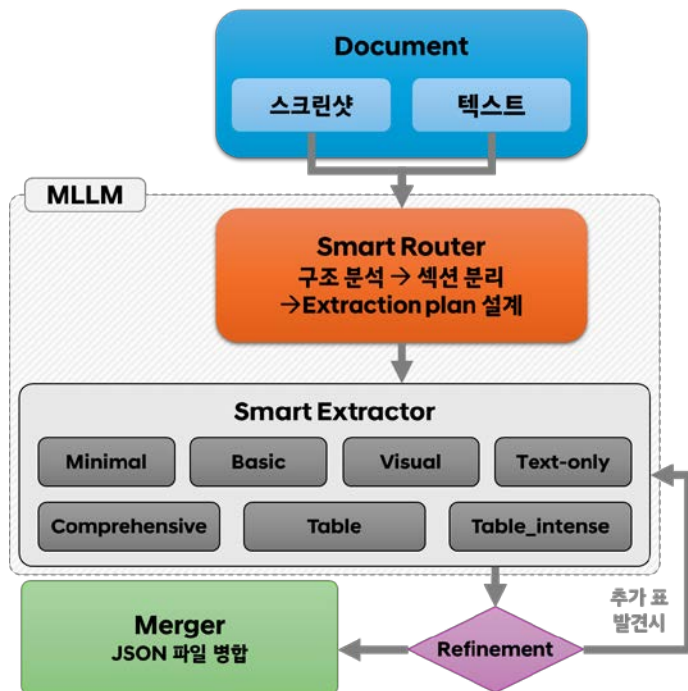


Figure 1 : An overview of semantic document arranger

제안하는 방법론은 멀티모달 LLM을 강력한 의미론적 배열기(semantic arranger)로 활용하는 모듈식 아키텍처로 구성되며, 적응형 정제 루프(adaptive refinement loop)를 통해 LLM이 추출한 텍스트를 검증·보정한다(Fig. 1). OCR과 문서 구조 분석을 위한 복잡한 보조 모델에 의존하지 않고, 페이지 스크린샷과 원시 텍스트를 함께 LLM에 입력함으로써 “문서 구조를 보면서 동시에 내용을 읽는” 총체적 이해를 수행하도록 설계했다.

각 페이지의 처리흐름은 다음과 같은 다단계 적응형 프로세스로 진행된다. 먼저 페이지별 스크린샷과 텍스트가 라우터(Router)로 전달되어 초기 레이아웃을 신속하고 포괄적으로 스캔한다(Fig. 1). 라우터는 제목, 요약, 문단을 포함한 문서의 개략적 초기 구조를 생성하고, 섹션 단위로 7개 카테고리(예: text only, table, visual 등)로 분류한다. 본 연구에서는 다양한 패턴을 예제로 다루도록 프롬프트를 설계하는 포괄적 퓨샷 프롬프팅(exhaustive few shot prompting) 기법을 통해 라우터 분류 정확도를 개선하였다.

다음으로 카테고리화된 각 섹션은 전문화된 추출기(extractor)에서 처리한다. 각 추출기는 템플릿 기반의 퓨샷 프롬프팅(template based few shot prompting)을 통해 각 유형에 맞는 JSON 객체로 데이터를 구조화한다. 예를 들어 “table”로 태깅된 섹션은 제공된 텍스트 배열을 그대로 사용하지 않고, 시각적으로 인식된 표에 맞게 “table_title”, “headers”, “rows” 스키마에 맞춰서 직렬화하도록 명시적으로 지시한다.

초기 JSON 객체 추출이 완료되면 정제 분석기(Refinement Analyzer)가 데이터가 정확하게 추출되었는지 여부를 점검한다. 추출된 텍스트에 휴리스틱(heuristics) 기법을 적용하여, 높은 수치 문자 밀도, 비슷한 길이 행의 반복, 일관된 공백 혹은 구분자의 존재 확인을 통해 추출된 데이터에서 누락된 표의 존재 확인 후 필요시 추출기에서 재처리 작업을 진행한다. 이러한 과정을 통해 추출된 데이터는 페이지 별로 병합되어 다양한 메타데이터들과 함께 최종 JSON 형태로 산출된다.

제안 기법의 유효성 검증을 위해 [Alphabet 25년 1분기 실적 보고서](#) (영문) 및 [현대자동차 아이오닉9 오너스 매뉴얼](#)(한글)을 테스트 코퍼스로 사용하였다. 대조군은 표준 텍스트 추출로 구성된 벡터 저장소를 생성하였고, 실험군은 제안된 파이프라인을 통해 생성된 직렬화된 JSON 출력으로 구성된 벡터 저장소를 생성하였다.

[사용 모델] 라우터 : Claude Opus 4.1 (높은 추론 능력), 추출기 : Gemini 2.5 Pro (높은 이미지 프로세싱 능력).

동일 조건 실험을 위해 각 문서의 표 내에서 답변 가능한 20개의 구체적 정량질문(예: “2024년 1분기 동안 Alphabet이 연구개발에 지출한 금액은 얼마입니까?”)을 구상하였다. 실험 시스템은 동일한 RAG 아키텍처를 유지하면서 실험/대조 벡터 저장소에 20개의 질문을 각각 질문하여 결과를 Table 1에 나타내었다.

비고	대조군 (Baseline)	실험군 (제안 파이프라인)
ALPHABET Q1 2025 실적 보고서 (영문)		
- 검색 정밀도 (Retrieval Precision) / %	75%	95%
- 사실 정확도 (Factual Accuracy, 2점 만점)	1.3 / 2.0	1.8 / 2.0
아이오닉9 오너스 매뉴얼 (국문)		
- 검색 정밀도 (Retrieval Precision) / %	65%	90%
- 사실 정확도 (Factual Accuracy, 2점 만점)	1.0 / 2.0	1.7 / 2.0

Table.1 Retrieval precision & Factual accuracy comparison for baseline and proposed pipeline (in en/ko docs.)

대조군은 올바른 문맥을 검색하는 데 어려움을 겪어 검색 정밀도(retrieval precision) 평균 70%를 기록하였고, 사실정확도(factual accuracy)는 1.15/2.0를 받았다. 반면 제안 시스템은 검색 정밀도 92.5%, 생성 응답의 사실 정확도는 1.75으로 크게 향상되었다. 제안된 파이프라인에서 생성된 메타데이터가 검색 정밀도에, 직렬화된 데이터가 사실 정확도에 기여한 것으로 보인다. 또한 정성적 비교에서도 뚜렷한 차이를 보였는데, 대조군 시스템은 빈번히 답변을 실패하거나 부정확한 수치 보인 반면, 실험군은 검색된 구조화 데이터에 기반하여 일관되게 정확한 답변을 제공하고 일부 상용 시스템보다 정확한 성능을 보이기도 했다.

본 연구는 시각적 데이터 구조를 보존하는 직렬화가 데이터 밀집 도메인에서 RAG 시스템의 신뢰성을 개선하는 핵심 전처리 단계임을 확인하였다. 제안하는 2단계 파이프 라인 2단계 파이프라인은 표·차트의 레이아웃을 해석해 서술적이고 의미 보존적인 JSON으로 변환함으로써, 텍스트 중심 RAG에서의 색인·질의 응답 성능을 향상시키는 동시에 별도의 VLM/OCR 구성 없이 운영 복잡도를 낮춘다. 향후 연구 방향으로는 (i) 라우터·추출기 역할에 사용하는 LLM 선택이 데이터 품질에 미치는 영향, (ii) 비정형 문서로부터 대규모 지식 그래프(knowledge Graph)로의 확장 가능성 평가 등 분야에서 추가적인 연구·검토가 가능할 것이다. 이상의 멀티모달 LLM 기반 직렬화 접근법은 보다 포괄적이고 신뢰성 높은 문서 데이터베이스 시스템 구축을 위해 기여할 것을 기대한다.