www.ksii.or.kr

2026년
추계학술 발표대회

10월 22일(목) – 10월 24일(토)

세인트존스호텔 (강원도 강릉시)

주최 · 주관

사단법인 한국인터넷정보학회

후 원

㈜나루씨큐리티, ㈜피씨엔, ㈜아이티센엔텍, 한화시스템㈜,
㈜LIG시스템, 케이투웹테크㈜, 강원관광재단, 강릉관광개발공사



www.ksii.or.kr
KSII
사단법인 KOREAN SOCIETY FOR INTERNET INFORMATION 한국인터넷정보학회

논문발표

10월 23일(금) 10:40~12:10 (Room 5 / 바부다 4)

세션 I	좌장: 김영철 교수(홍익대)

AI-Metaverse

I-1 가상 융합 쇼핑몰에서의 테스트 에이전트 기반 자동 테스트 메커니즘
정현욱, 김장환, 김영철(홍익대)

I-2 가상 융합 전시관 내 시각·공간 문맥 활용 기반 고도화 RAG 지향 AI 도슨트 구현 사례
김인용, 김장환, 서채연(홍익대), 전병국(강원대), 김영철(홍익대)

I-3 사용자 의도와 생성형 AI 결과의 일치도 평가를 위한 개념 정의 및 평가방법
정연주, 서채연, 김영철(홍익대)

I-4 사용자 선호도 메트릭스 기반 Text-to-Video 생성 결과 평가 방안
정연주, 서채연, 김영철(홍익대)

I-5 LLM기반 AI챗봇 자가 치유 메커니즘
김성회, 서채연, 김영철(홍익대)

10월 23일(금) 10:40~12:10 (Room 6 / 카리브 2)

[워크숍] XR 콘텐츠 자율변형 프레임워크 기술 세션 3	자체 진행

10월 23일(금) 10:40~12:10 (Room 7 / 카리브 1)

육사 AI R&D센터 특별섹션 세미나 1	좌장: 이준 교수(육사)

가상 융합 전시관 내 시각·공간 문맥 활용 기반 고도화 RAG 지향 AI 도슨트 구현 사례

김인용[1], 김장환[2], 서채연[3], 전병국[4], 김영철[5*]
홍익대학교 소프트웨어융합학과[1,2,3,5*], 강원대학교 컴퓨터공학과[4]
e-mail : citefred@gmail.com[1], {lentoconstante[2], chaeyun[3], bob[5*]}@hongik.ac.kr,
jeonbk@gwnu.ac.kr[4]

Implemented Practice on Advanced RAG-driven AI Docent Based on utilizing Visual and Spatial Context in Virtual Exhibitions

Inyong Kim[1], Janghwan Kim[2], Chaeyun Seo[3], Byungkook Jeon[4], R. Young Chul Kim[5*]

1. 서 론

가상현실(VR) 전시는 시공간적 제약을 극복하는 대안으로 주목받고 있으나, 대다수 일회성 하드코딩 방식으로 제작되어 유지보수와 재사용에 어려움이 있다. 이를 해결하고자 선행 연구에서는 웹 CMS와 건축 도면(BIM)을 연동하여 시각적 재사용성을 지닌 가상 전시관을 구축하고, RAG 파이프라인을 결합해 AI 도슨트의 지식 재사용성을 구현했다[1][2][3].

그러나 기존 로컬 문서 기반 RAG는 외부 지식 확장에 한계가 있다. 또한 텍스트 중심 처리로 인해 작품의 시각적 특징이나 관람객의 좌표, 시선 등 공간 맥락(Spatial Context)을 상호작용에 반영하지 못하는 단방향적 한계를 지닌다.

이를 해결하고자 본 연구는 웹 검색(동적 지식 확장), 멀티모달(시각 정보 인지), 가상 3D 공간 위치 정보(좌표 및 시선 기반 상호작용)를 통합하여 기존 RAG를 고도화한 AI 도슨트 파이프라인 구현 사례를 제안한다. 본 연구는 2장 관련 연구, 3장 설계 및 구현, 4장 결론으로 구성된다.

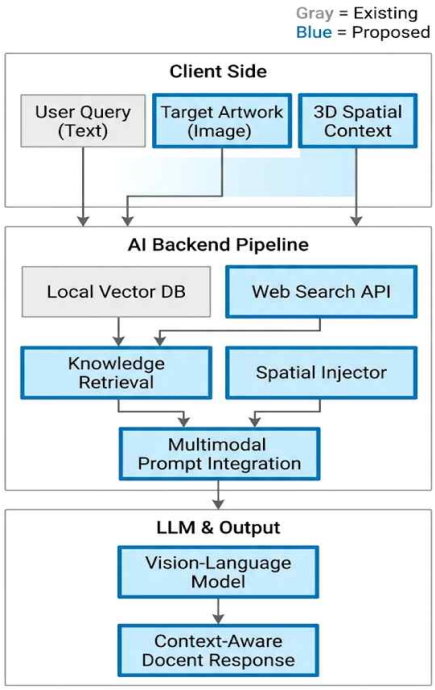
2. 관련 연구

가상 전시관의 지속 가능성을 확보하기 위해, 선행 연구에서는 웹 CMS와 검색 증강 생성(RAG) 파이프라인을 결합하여 데이터와 지식의 재사용성을 구현하였다[1][2][3]. RAG 기술은 모델 외부의 지식을 동적으로 주입하여 대형언어모델(LLM)의 환각(Hallucination) 현상을 제어하는 표준으로 활용된다[4]. 그러나 선행 연구의 시스템은 사전에 구축된 로컬 문서에만 의존하여 최신 전시 동향 등 실시간 지식 반영이 어렵다. 정적 언어 모델의 한계를 극복하기 위해 검색 엔진을 융합하여 외부 지식을 동적으로 수용하는 연구가 입증된 바 있으며[5], 가상 전시 환경의 AI 도슨트 역시 단순 내부 검색을 넘어 웹 기반 지식 확장이 필수적이다.

나아가 전시관은 시각 예술 공간임에도 기존 시스템은 텍스트 중심의 단방향 처리에 머물러 있다. 최근 텍스트와 이미지를 동시에 검색하고 융합하는 멀티모달 RAG 아키텍처가 제안되어 시각적 맥락 파악 성능의 향상을 입증하였다[6]. 또한, 3D 가상 환경 내에서 에이전트와 사용자의 물리적 좌표 및 시야 정보를 결합한 질의응답(Embodied QA) 모델은 현장감과 상호작용 수준을 극대화하는 것으로 확인되었다[7]. 따라서 본 연구는 선행 연구의 한계를 보완하고, 웹 검색(동적 지식 확장)[5], 멀티모달(시각 인지)[6], 가상 3D 공간 위치 정보[7]를 통합 활용하는 고도화된 AI 도슨트 파이프라인을 제안한다.

3. 멀티 모달 AI 도슨트 파이프라인 설계 및 구현

전시관 클라이언트의 시·공간 및 질의 데이터가 AI 백엔드의 검색 파이프라인과 결합해 최종 해설로 생성되는 흐름을 구현하고자 한다. 제안하는 웹 검색, 멀티모달 프롬프트, 공간 맥락 동기화 모듈이 결합된 AI 도슨트 파이프라인의 설계 및 구현 시스템의 구조도는 그림 1과 같다.

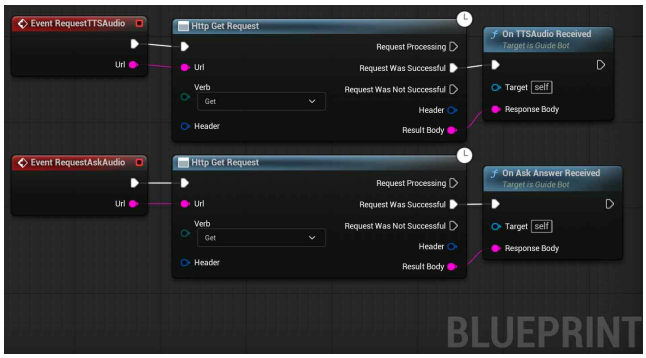


(그림 1) 시스템 전체 구성도

3.1 동적 지식 확장을 위한 웹 기반 RAG 구현

기존 정적 Vector DB의 한계를 극복하기 위해 사용자 질의를 분석하여 로컬 검색과 웹 검색을 동적으로 라우팅하는 검색 파이프라인을 구현하였다. 작품의 기본적인 메타데이터와 도록 설명은 CMS와 연동된 내부 Vector DB에서 검색(Retrieval)한다. 반면, 최신 전시 동향이나 외부 평가 등 확장된 지식이 요구되는 질의는 웹 검색 API를 호출하도록 설계하였다. 추출된 내·외부 텍스트 데이터는 최종 프롬프트의 컨텍스트(Context) 영역에 병합되어 주입되며, 이를 통해 파인튜닝 없이 실시간 지식이 반영된 해설 생성이 가능하다. 그림

2는 관람객의 질의 데이터를 백엔드 서버로 요청하는 언리얼 엔진의 비주얼 스크립트 인터페이스의 모습이다.



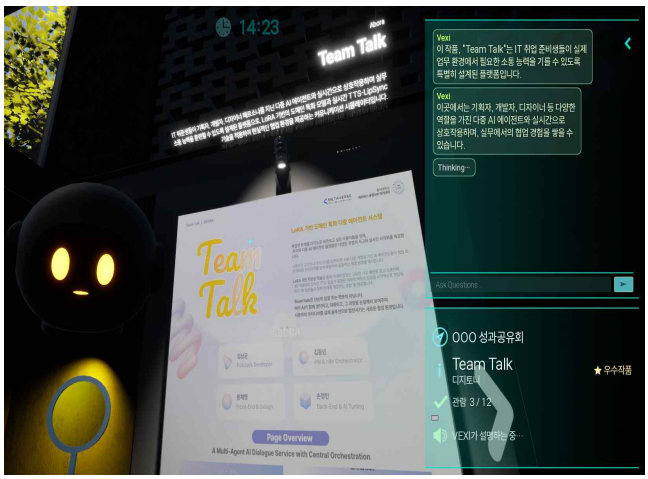
(그림 2) Unreal Engine의 AI 요청부 인터페이스

3.2 시각 정보 처리를 위한 멀티모달 프롬프트 구현

작품의 시각적 특징을 AI가 직접 인지하고 해설에 반영할 수 있도록 멀티모달(Multimodal) 처리 구조를 구현하였다. 관람객이 시각적 요소를 포함한 질문을 할 경우, 시스템은 CMS에 등록된 원본 작품 이미지 또는 사용자의 현재 뷰포트 화면을 캡처하여 시각-언어 모델(Vision-Language Model)의 입력값으로 전달한다. 텍스트 지식과 이미지 데이터가 통합된 멀티모달 프롬프트를 통해, AI 도슨트는 단순 텍스트 매칭을 넘어 색채, 구도 등의 시각적 맥락을 바탕으로 교차 추론된 답변을 생성한다.

3.3 가상 3D 공간 맥락 동기화 및 시스템 통합

가상 전시관 내 관람객의 몰입감을 높이기 위해 엔진 상의 물리적 상태 정보를 도슨트 시스템에 연동하였다. 관람객 아바타의 3D 좌표(X, Y, Z)와 시선 벡터(Forward Vector) 데이터를 실시간으로 추출하여 프롬프트의 메타데이터로 주입하는 모듈을 구현하였다. 이를 통해 AI 도슨트는 관람객과 특정 작품 간의 거리 및 주시 방향을 인지하여 공간 기반 질의(예: "내 오른쪽에 있는 작품 설명해 줘")에 대한 맥락 맞춤형 응답이 가능하다. 최종적으로 지식 검색, 시각 인지, 공간 맥락 모듈이 하나의 파이프라인으로 통합되어 상황 인식형 도슨트 상호작용을 완성한다. 그림 3은 파이프라인에서 생성된 AI 도슨트 대본이 가상 공간 내 3D 아바타의 UI를 통해 관람객에게 출력되는 실시간 상호작용 화면이다.



(그림 3) AI 파이프라인의 클라이언트 도슨트 아바타

4. 결 론

본 연구는 웹 검색 기반의 동적 지식 확장, 멀티모달 프롬프트를 통한 시각 정보 인지, 그리고 가상 3D 공간 맥락 데이터를 통합하여 고도화된 멀티 모달 AI 도슨트 파이프라인을 설계 및 구현하였다.

이를 통해 기존 정적 로컬 문서 기반 RAG의 단방향적 한계를 극복하고, 관람객의 시야와 위치, 최신 외부 지식까지 종합적으로 반영하는 현장감 높은 상호작용을 파인튜닝(Fine-tuning) 없이 성공적으로 구현하였다.

향후 연구에서는 멀티모달 및 웹 검색 연동으로 인해 추가로 발생하는 응답 지연 시간(Latency)을 효과적으로 은닉하기 위한 비동기 UX 마스킹 기법을 적용할 예정이다. 나아가 AI가 관람객의 동선 및 시선 체류 데이터를 분석하여 가상 공간의 테마를 동적으로 변환하는 지능형 메타버스 프레임워크로 본 연구를 확장할 계획이다.

Acknowledgment

본 연구는 2025년도 문화체육 관광부의 재원으로 한국콘텐츠진흥원(과제명: 인공지능 기반 대화형 멀티모달 인터렉티브 스토리텔링 3D장면 저작 기술 개발, 과제번호:RS-2023-00227917, 기여율:100%) 지원과 한국연구재단의 4단계 두뇌한국21사업(과제명: 초분산 자율 컴퓨팅 서비스 기술 연구팀, 과제번호:202003520007)의 지원을 받아 수행된 연구임. 또한, 홍익대학교 메타버스 융합 SW 아카데미 7기 교육생들의 프로젝트 결과물이다.

참 고 문 헌

[1] 김인용, 조형구, 마예은, 오재은, 공지훈, 김장환, 김영철, "건축 도면 기반 VR 가상 전시관 개발 모범 사례," 2025 한국인터넷정보학회 추계학술발표대회, 2025.

[2] 김인용, 조형구, 마예은, 공지훈, 김기두, 김장환, 김영철, "통합 콘텐츠 관리 시스템 기반 재사용 VR 전시관 구현 사례," 2025 한국스마트미디어학회 추계학술대회, 2025.

[3] 김인용, 조형구, 공지훈, 김인권, 김장환, 서채연, 김영철, "가상 전시관 내 RAG 기반 AI 도슨트의 데이터 및 지식 재사용성 구현 사례," 한국인터넷정보학회 춘계학술발표대회, 2026.

[4] P. Lewis 외, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Proc. of Advances in Neural Information Processing Systems (NeurIPS), 2020.

[5] T. Vu 외, "FreshLLMs: Refreshing Large Language Models with Search Engine Augmentation," in Proc. of the Association for Computational Linguistics (ACL), 2024.

[6] W. Chen 외, "MuRAG: Multimodal Retrieval-Augmented Generator for Open Question Answering over Images and Text," in Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.

[7] A. Das 외, "Embodied Question Answering," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.