

국립중앙박물관 Retriever AI Docent Chatbot

슈팅스타 With 셀렉트스타

주성민, 박상민, 이성준, 최우성, 최효정



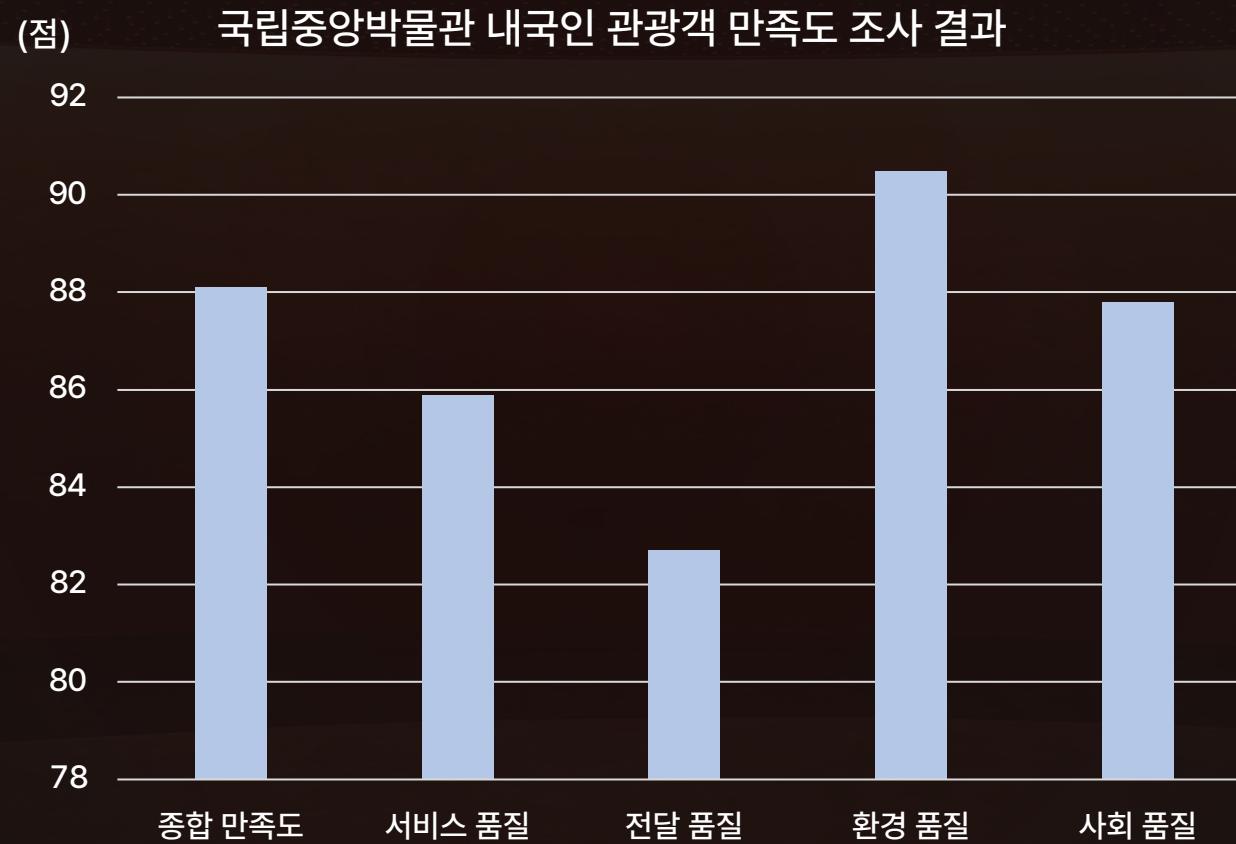
Index

- 프로젝트 배경 3p
- LLM의 한계 8p
- RAG 모델 11p
- Retriever 고도화 (DPR) 29p
- 프로토타입 구현 이미지 35p
- 성능 비교 38p
- 결론 및 한계점 43p

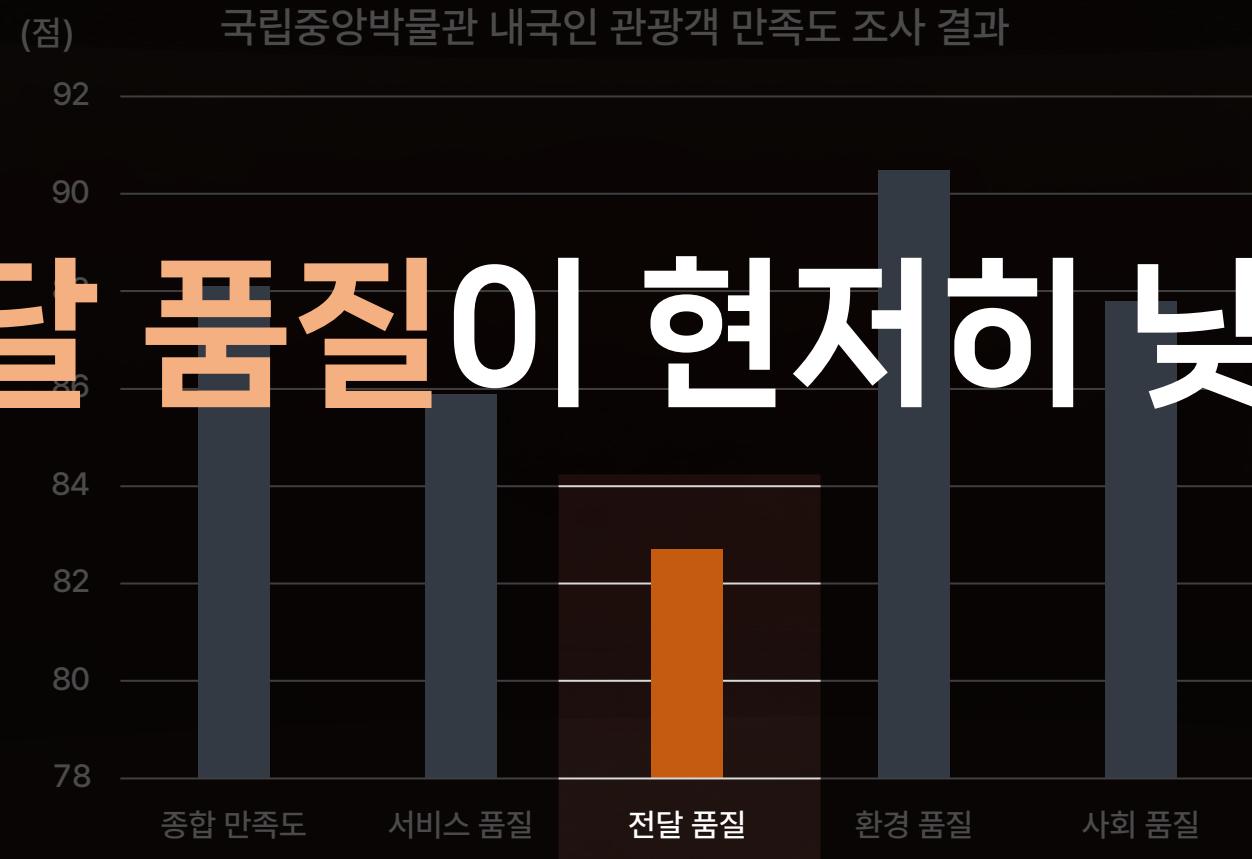


프로젝트 배경

출처: 국립중앙박물관 고객만족도 조사보고서 (2022)



출처: 국립중앙박물관 고객만족도 조사보고서 (2022)

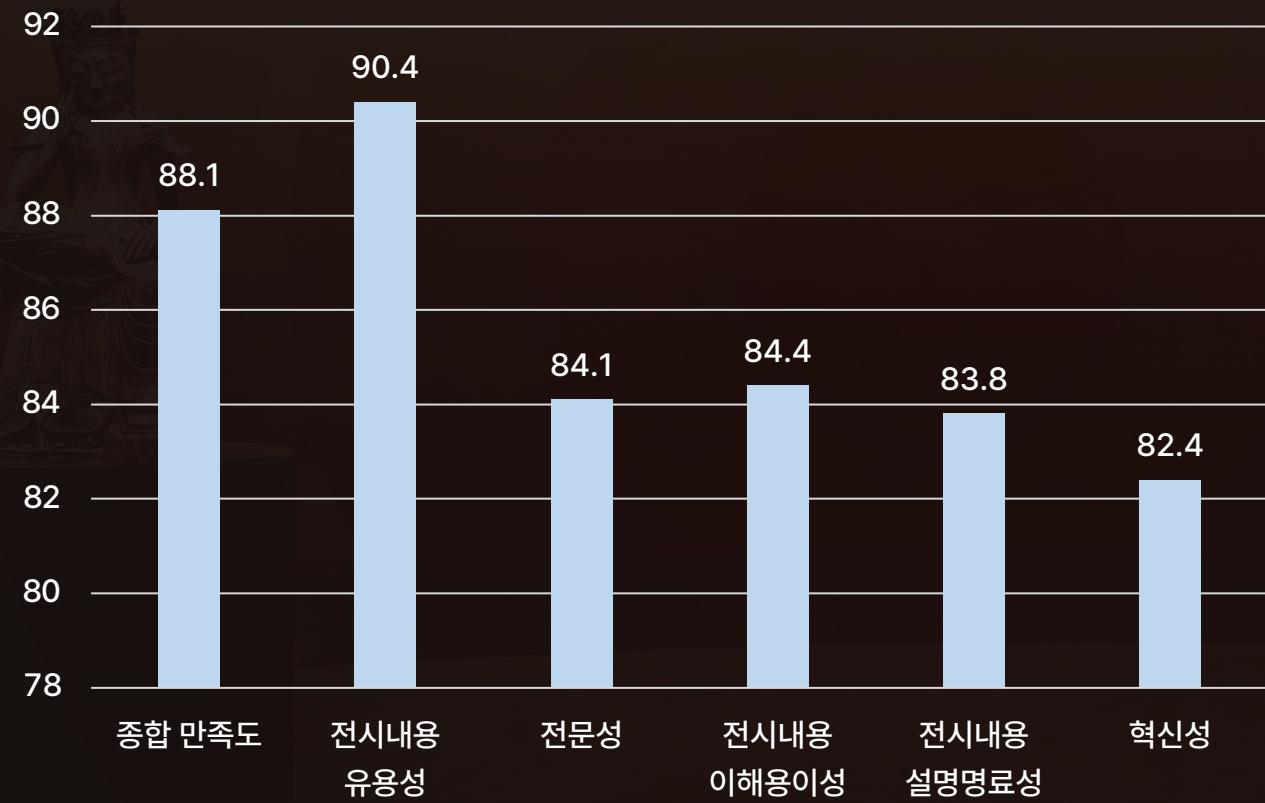


전달 품질이 현저히 낮음

프로젝트 배경

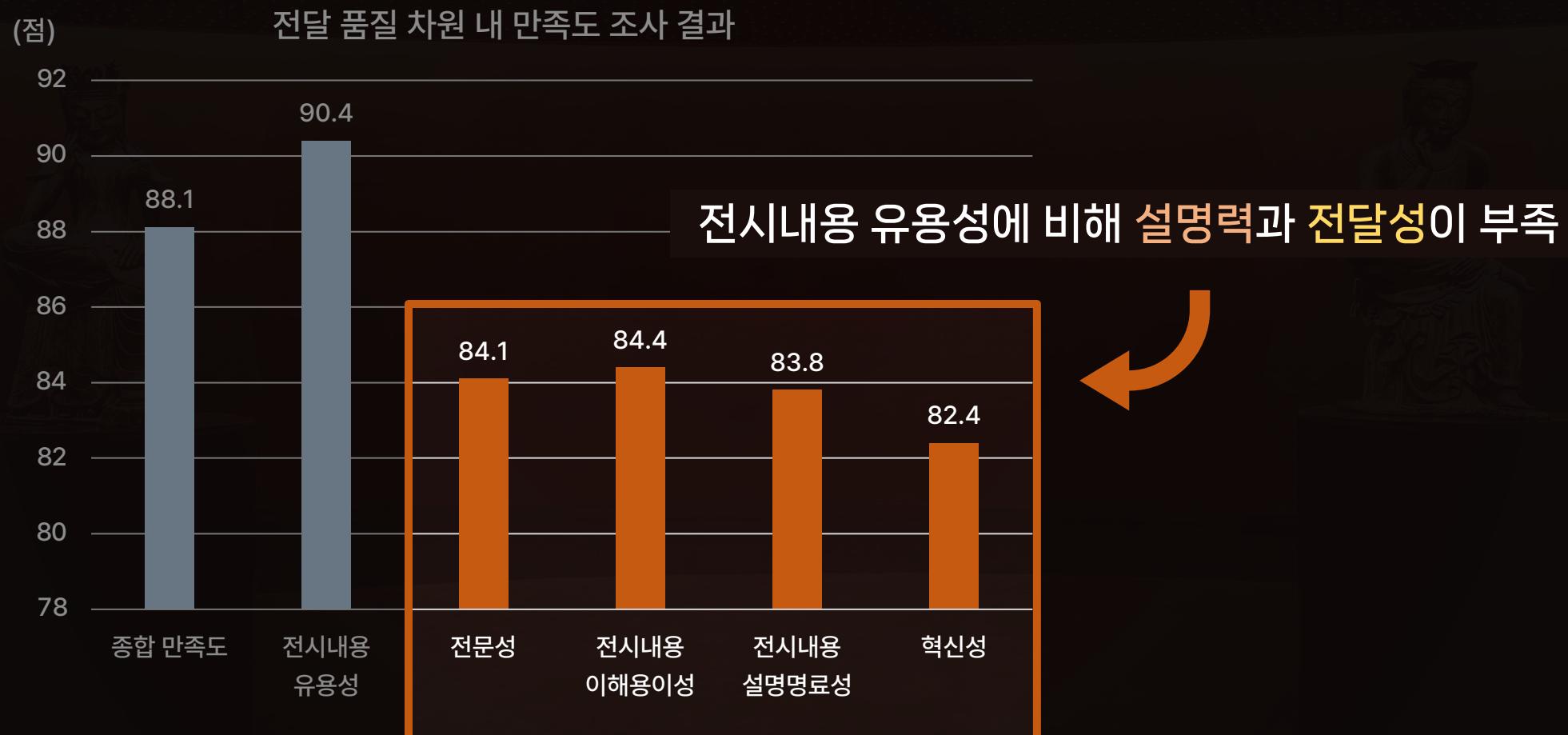
출처: 국립중앙박물관 고객만족도 조사보고서 (2022)

(점) 전달 품질 차원 내 만족도 조사 결과



프로젝트 배경

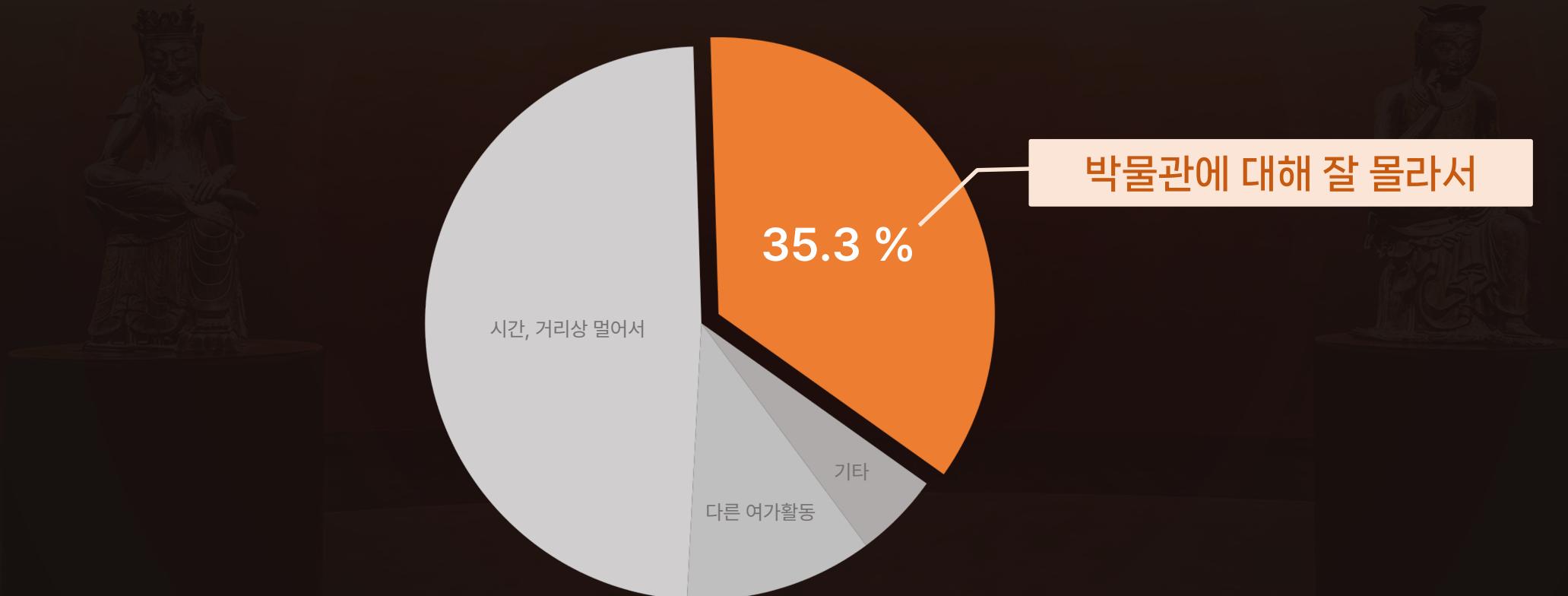
출처: 국립중앙박물관 고객만족도 조사보고서 (2022)



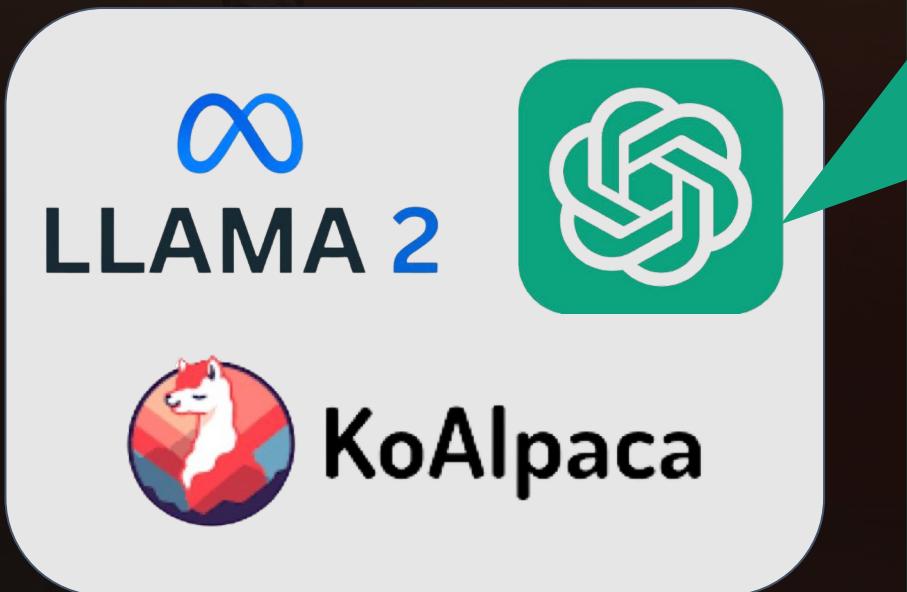
프로젝트 배경

국립중앙박물관에 방문하지 않는 이유는?

출처: 국립중앙박물관 고객만족도 조사보고서 (2022)



LLM의 한계점



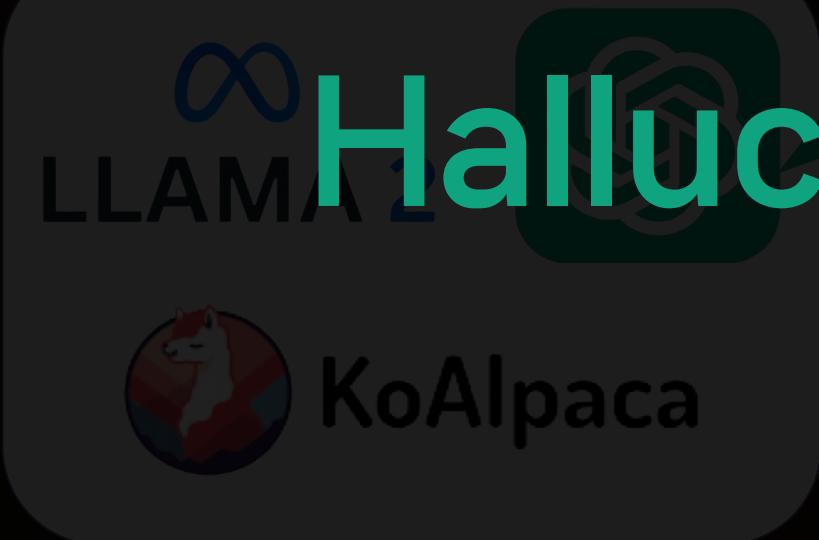
효정 You
이어드림스쿨에 대해 알려줘

ChatGPT
이어드림스쿨은 한국어로 "이포아드"라고도 불리며, 한국에서 인기 있는 인터넷 연재 웹툰 플랫폼 중 하나입니다. 이어드림스쿨은 다양한 장르의 웹툰을 연재하고, 사용자들은 이 플랫폼을 통해 웹툰을 읽을 수 있습니다.

주로 웹툰 작가들이 자신의 작품을 연재하고 독자와 소통하며 웹툰을 공개하는 데 사용하는 플랫폼 중 하나로 알려져 있습니다. 이어드림스쿨에서는 다양한 스토리와 그림체의 웹툰을 찾아볼 수 있으며, 사용자들은 웹툰을 구독하고 좋아하는 작품을 읽을 수 있습니다.

이어드림스쿨은 웹툰 팬들 사이에서 인기가 있으며, 다양한 웹툰 작품을 즐기고 공유할 수 있는 플랫폼으로 활용되고 있습니다. 이와 관련된 더 자세한 정보는 이어드림스쿨의 공식 웹사이트나 앱을 통해 얻을 수 있습니다.

LLM의 한계점



이어드림스쿨에 대해 알려줘

ChatGPT
이어드림스쿨은 한국어로 "이포아드"라고도 불리운 사이트에서 운영하는 인터넷 연재 웹툰 플랫폼 중 하나입니다. 이어드림스쿨은 다양한 장르의 웹툰을 업로드하고 독자들이 이를 플랫폼을 통해 웹툰을 읽을 수 있습니다.

Hallucination (환각)

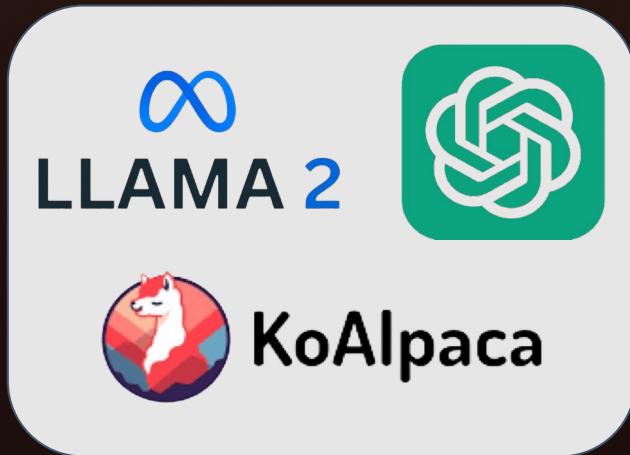
주로 웹툰 작가들이 자신의 작품을 연재하고 독자와 소통하며 웹툰을 공개하는 데 사용하는 플랫폼 중 하나로 알려져 있습니다. 이어드림스쿨에서는 다양한 스토리와 그림체의 웹툰을 찾아볼 수 있으며, 사용자들은 웹툰을 구독하고 좋아하는 작품을 읽을 수 있습니다.

이어드림스쿨은 웹툰 팬들 사이에서 인기가 있으며, 다양한 웹툰 작품을 즐기고 공유할 수 있는 플랫폼으로 활용되고 있습니다. 이와 관련된 더 자세한 정보는 이어드림스쿨의 공식 웹사이트나 앱을 통해 얻을 수 있습니다.

LLM Fine-Tuning의 한계점

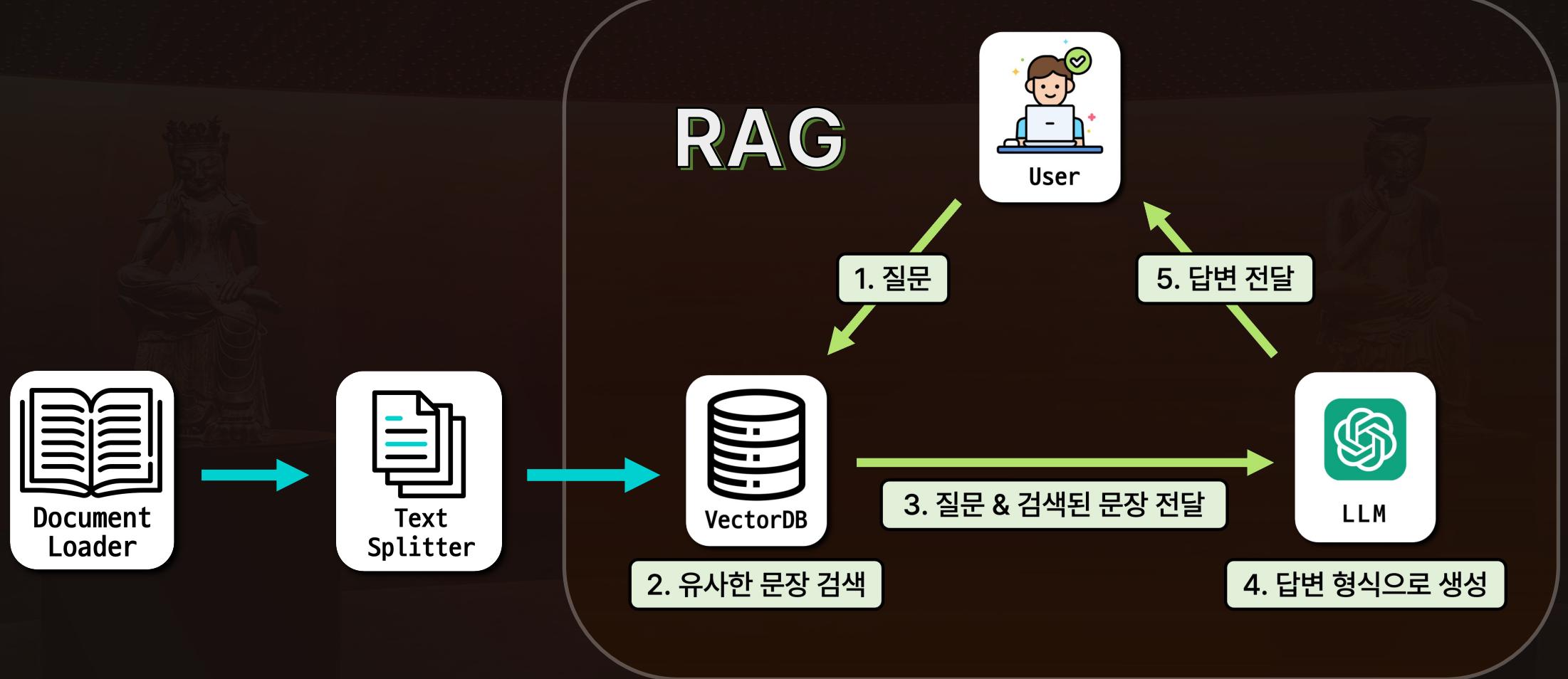


비용과 자원

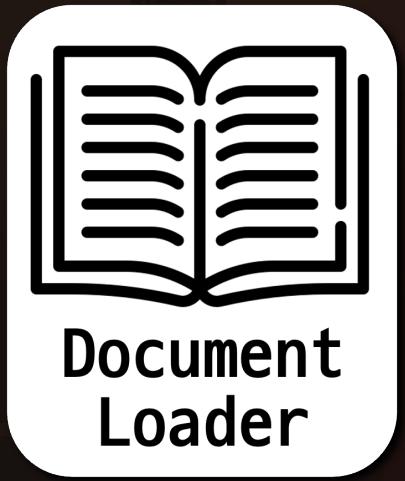


새로운 정보 반영 한계

RAG 모델 소개



데이터 수집 방식

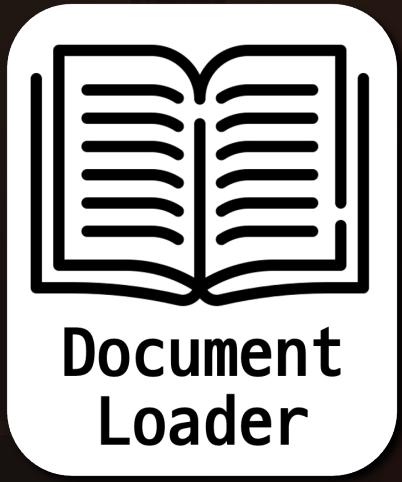


크롤링

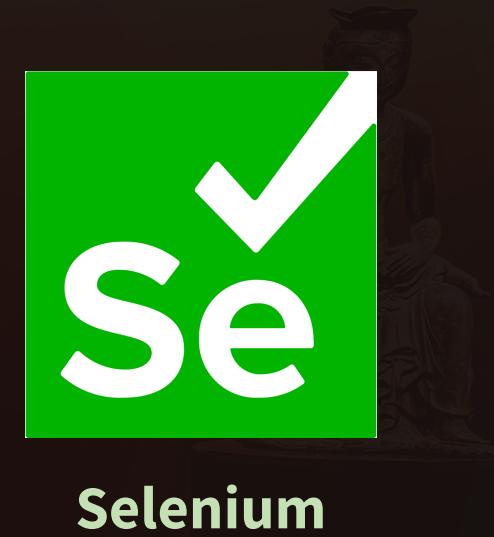
데이터 수집 방법



데이터 수집 방식

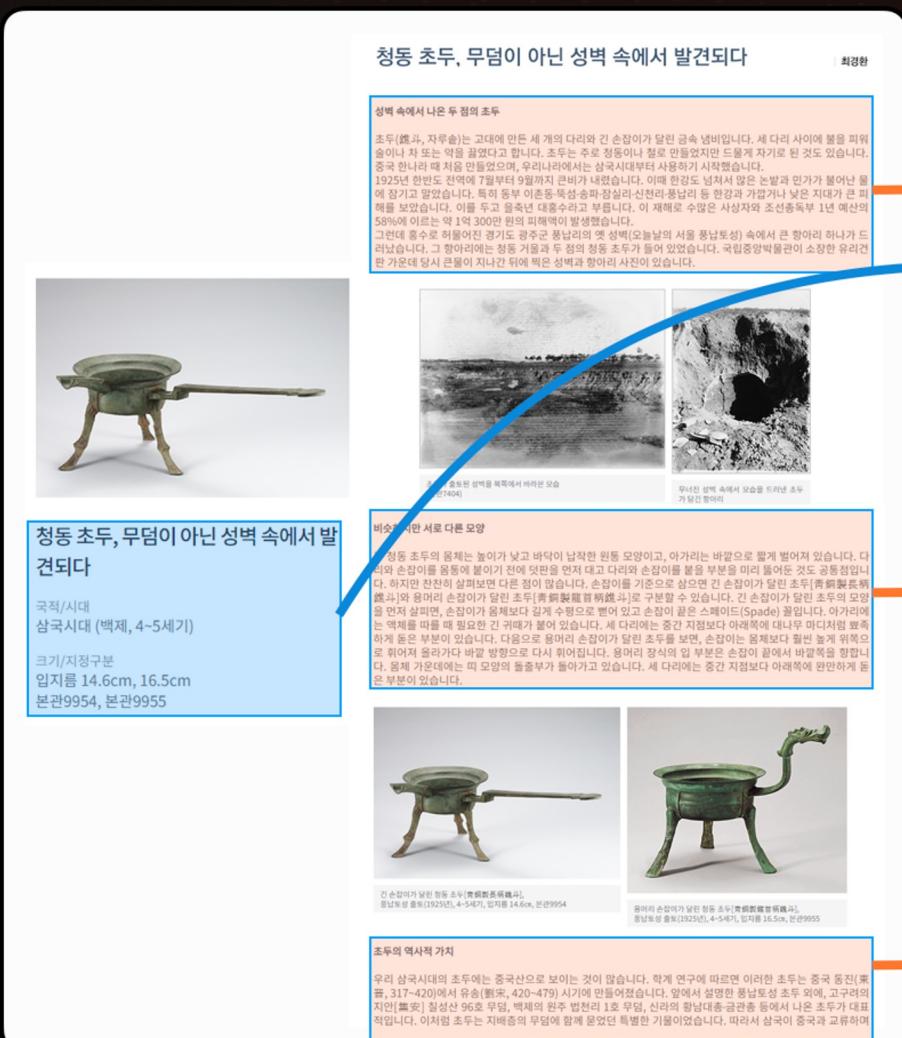


BeautifulSoup +
BeautifulSoup4



데이터 수집 방법

데이터 수집 및 구조화 방식



1 {"title": "청동 초두, 무덤이 아닌 성벽 속에서 발견된다", "link": "https://www.museum.go.kr/site/main/relic/recommend/view?relicRecommenId=1569309", "country_and_era": "삼국시대 (백제, 4~5세기)", "size_and_category": "입지름 14.6cm, 16.5cm\n본관9954, 본관9955", "description": "성벽 속에서 나온 두 점의 초두[鎚斗, 자루술]는 고대에 만든 세 개의 다리와 긴 손잡이가 달린 금속 술이나 차 또는 약을 끓였다고 합니다. 초두는 주로 청동이나 철로 만들었지만 드물게 자기로 된 것도 있습니다. 중국 한나라 때 처음 만들어졌습니다. 우리나라에서는 삼국시대부터 사용하기 시작했습니다. 1925년 한반도 전역에 7월부터 9월까지 큰비가 내렸습니다. 이때 한강도 넘쳐서 많은 논밭과 민가가 불어난 물에 잠기고 말았습니다. 특히 동부 이촌동·뚝섬·송파·잠실리·신천리·풍납리 등 한강과 가깝거나 낮은 지대가 큰 피해를 보았습니다. 이를 두고 을축년 대홍수라고 부릅니다. 이 재해로 수많은 사상자와 조선총독부 1년 예산의 58%에 이르는 약 1억 300만 원의 피해액이 발생했습니다. 그런데 홍수로 허물어진 경기도 광주군 풍납리의 옛 성벽(오늘날의 서울 풍납토성) 속에서 큰 항아리 하나가 드러났습니다. 그 항아리에는 청동 거울과 두 점의 청동 초두가 들어 있었습니다. 국립중앙박물관이 소장한 유리건판 가운데 당시 큰물이 지나간 뒤에 찍은 성벽과 항아리 사진이 있습니다. \n초두가 출토된 성벽을 북쪽에서 바라본 모습(건판7404)\n무너진 성벽 속에서 모습을 드러낸 초두가 담긴 항아리(건판7406)"}

2 {"title": "청동 초두, 무덤이 아닌 성벽 속에서 발견된다", "link": "https://www.museum.go.kr/site/main/relic/recommend/view?relicRecommenId=1569309", "country_and_era": "삼국시대 (백제, 4~5세기)", "size_and_category": "입지름 14.6cm, 16.5cm\n본관9954, 본관9955", "description": "비슷하지만 서로 다른 모양은 높이가 낮고 바닥이 납작한 원통 모양이고, 아가리는 바깥으로 짧게 벌여져 있습니다. 다리와 손잡이를 몸통에 붙이기 전에 덧판을 먼저 대고 다리와 손잡이를 볼을 부분을 미리 뚫어둔 것도 공통점입니다. 하지만 찬찬히 살펴보면 다른 점이 많습니다. 손잡이를 중심으로 삼으면 긴 손잡이가 달린 초두[青銅製長柄鎚斗]와 용머리 손잡이가 달린 초두[青銅製龍首柄鎚斗]로 구분할 수 있습니다. 긴 손잡이가 달린 초두의 모양을 먼저 살펴보면, 손잡이가 움푹보다 아래쪽에 대나무 마디처럼 뾰족하게 돋은 부분이 있습니다. 다음으로 용머리 손잡이가 달린 초두를 보면, 손잡이는 몸통보다 훨씬 위쪽으로 휘어져 물리거나 대각 방향으로 다시 휘어집니다. 용머리 장식의 입 부분은 손잡이 끝에서 바깥을 향합니다. 몸체 가운데에는 띠 모양의 돌출부가 돌아가고 있습니다. 세 다리에는 중간 지점보다 아래쪽에 완만하게 돋은 부분이 있습니다."}

3 {"title": "청동 초두, 무덤이 아닌 성벽 속에서 발견된다", "link": "https://www.museum.go.kr/site/main/relic/recommend/view?relicRecommenId=1569309", "country_and_era": "삼국시대 (백제, 4~5세기)", "size_and_category": "입지름 14.6cm, 16.5cm\n본관9954, 본관9955", "description": "초두의 역사적 가치[우리 삼국시대의 초두에는 중국산으로 보이는 것이 없습니다. 학계 연구에 따르면 이러한 초두는 중국 동진(東晉, 317~420)에서 유송(劉宋, 420~479) 시기에 만들어졌습니다. 앞에서 설명한 풍납토성 초두 외에, 고구려의 지안[集安] 칠성산 96호 무덤, 백제의 원주 법천리 1호 무덤, 신라의 황남대총·금관총 등에서 나온 초두가 대표적입니다. 이처럼 초두는 지배층의 무덤에 함께 묻었던 특별한 기물이었습니다. 따라서 삼국이 중국과 교류한 것입니다.]

데이터 수집 및 구조화 방식

<h5> + <p>를 한 문단으로 묶어서 미리 Chunking

h5 589 x 18.5

성벽 속에서 나온 두 점의 초두

p 589 x 260 두 점의 초두

초두(鎧斗, 자루솥)는 고대에 만든 세 개의 다리와 긴 손리 사이에 불을 피워 술이나 차 또는 약을 끓였다고 합니다. 하지만 드물게 자기로 된 것도 있습니다. 중국 한나라 삼국시대부터 사용하기 시작했습니다.

1925년 한반도 전역에 7월부터 9월까지 큰비가 내렸습니다. 민가가 불어난 물에 잠기고 말았습니다. 특히 동부 이리 등 한강과 가깝거나 낮은 지대가 큰 피해를 보았습니다. 이 재해로 수많은 사상자와 조선총독부 1년 예산의 피해액이 발생했습니다.

그런데 홍수로 허물어진 경기도 광주군 풍납리의 옛 성 큰 항아리 하나가 드러났습니다. 그 항아리에는 청동 거습니다. 국립중앙박물관이 소장한 유리건판 가운데 당시 사진이 있습니다.

<h5> 성벽 속에서 나온 두 점의 초두 </h5> == \$0

▶ <p> ... </p>

▶ <div class="thum ctr_v21_each cl_b" style="max-width: 590px; width: 100%; margin: 0px auto 20px;"> ... </div>

<h5> 비슷하지만 서로 다른 모양 </h5>

▶ <p> ... </p>

▶ <div class="thum ctr_v21_each cl_b" style="max-width: 726px; width: 100%; margin: 0px auto 20px;"> ... </div>

<h5> 초두의 역사적 가치 </h5>

▶ <p> ... </p>

<h5> 풍납토성 초두에 담긴 의미 </h5>

▶ <p> ... </p>

추후 Bert 계열 모델 적용에 대비함

데이터 구조화 방식

문제점 : 소제목 <h5>가 없거나 하나의 소단원이 너무 긴 경우 발생

큐레이터 추천 소장품 소장품 검색 국보·보물 검색 3D 데이터 검색 수어 동영상 음성 안내

황남대총은 1973년부터 1975년까지 문화재관리국 문화재연구소(현재의 문화재청 국립문화재연구소)에서 발굴하였습니다. 신라의 돌무지덧널무덤 가운데 규모가 가장 큰 것으로 속하며, 귀금속 장신구와 희귀한 수입품 그리고 엄청난 양의 철제품과 길그릇이 부장된 점에서 학자들은 이 무덤을 왕릉으로 판단하고 있습니다. 하지만 어느 왕의 능인지 정확히 알 수 없었으므로, 1976년에 문화재위원회는 이 무덤을 경주사 황남동에 있는 큰 무덤이란 뜻으로 '황남대총'이라는 이름을 붙였습니다. 황남대총은 남쪽으로 두 개의 무덤을 약간씩 차례로 끝을 통해 밟혀왔습니다. 남쪽과 북쪽의 무덤 모양은 돌무지덧널무덤이란 점에서 같으나, 나부구조에서 남쪽 무덤은 시선을 모신 주작(主釦)과 더불어 부장품을 가진 채로 부기(副基)를 떠로 둔 점에서 차이가 보입니다. 북쪽이 왕비의 능이었다는 결정적 증거로 은제 머리띠 꿈틀개가 세겨져 있었기 때문인데, 이 부인이란 표현이 당시 왕비를 말하는 것입니다. 즉 '夫人帶'은 왕비의 허리띠라는 의미를 지닌다고 할 수 있습니다. 황남대총 북분에 안장된 왕비는 금관과 금모로 꾸민 허리띠 이외에도 금팔찌, 금반지, 금걸개, 기술구미개 등으로 치장하였습니다. 비록 비단옷과 장신구에 대비되는 직물이 거의 모두 써어버려졌으나, 남아있는 귀한 장신구는 생전에 누렸던 가장 화려한 복장을 입혀 장례가 이루어졌음을 짐작할 수 있습니다.

왕비의 금관은 신라에서 유행한 나뭇가지모양 금관의 전형을 보여주고 있습니다. 세 개의 맞가지[互生枝]와 두 개의 엇가지[互生枝]로 조합된 세울장식[立輪飾]을 높이 솟은 오르도록 부착되었습니다. 맞가지는 정면과 좌우측에 세웠고, 엇가지는 위쪽으로 비스듬하도록 후면과 좌우측에 붙였습니다. 표면에는 무늬를 새기고 금은목과 달개를 충실히 해석하여 꾸몄습니다. 비취라고 부른 경옥으로 만든 금관을 금판을 통하게 오른 달개는 하나하나씩 금실을 매달았습니다. 머리띠 아래에 금은 고리에 사슬로 달개를 엮은 금제 드리개를 좌우에 세 개씩 늘어뜨렸습니다. 머리띠는 유기질로 된 것으로 앙장을 묶었던 듯 합니다. 굳은 고리의 금제 드리개는 적용한 사람의 여성성이었을 수 있음을 상정하고 있습니다. 이에 비해 대체로 남성의 귀걸이와 드리개는 거의 인경이나 경기가 많습니다. 이 금관은 신라의 나뭇가지 모양 금관에서 맞가지와 엇가지가 조합된 최초의 것입니다.

황남대총에서 가장 의아한 수수께끼는 왕과 왕비가 지녔던 관(眥)을 들러싼 문제입니다. 일반적인 생각과 전혀 다르게, 왕은 금관을 지냈고 왕비는 금관을 지나고 있었습니다. 구관판에 도금을 한 금동관과 순수한 금관은 재질에서 엄청난 차이를 보입니다. 비록 왕의 관은 유일하게 금은목을 단 금동관임에도 불구하고, 재질이 금동의 점에서 금관에 비해 결코 우월하다고 볼 수 없습니다. 뿐만 아니라 왕의 금동관은 세울장식이 세 개의 맞가지 뿐입니다. 이러한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다.

신라의 '나뭇가지모양 관'은 세울 장식의 모양에 따라 변화를 3단계로 나눌 수 있습니다. 처음에는 머리띠에 세 개의 맞가지만은 아니라 세 개의 엇가지로 '나뭇가지'로 알려온 자영스카[나뭇가지]와 봄날[봄날]과 봄날[봄날]과 봄날[봄날]로 변형됩니다. 이로써 신라의 왕에게는 전통제의와 관련된 황금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지나고 있던 제의적 상징은 그대로 민간에 전승되어 정치적 권위가 사라진 채 무당의 신물로 남게 되었다고 볼 수 있습니다.

왕비의 금관. 경주 황남대총 북분 출토, 높이 27.3 cm.
국보

나눌 수 있습니다. 처음에는 머리띠에 세 가지가 바로 벌어진 것에서 점차 직각 형태이므로 '사원형식(始原型式)'으로 분류는 황남대총 남분에서 왕이 착용한 금관을 있어 뒤이은 형식과는 이어집니다. '나뭇 것인데, 가장 다듬어진 형태이므로' 표준화되고 있습니다. 엊가지는 뒤늦게 더해졌으니 즉면서서 나무와 사슬들은 하늘과 땅을 걸치기 모두 직각으로 꺾인 형태입니다. 1) 및 서봉총(瑞鳳塚)의 금관과 더불어 표지의 마디가 3단인 점이 특징입니다. 이 전단계에 속하며, 엊가지의 마디가 4단인 3단인 것에서 4단인 것으로 바뀌었고, 전합니다. 마지막인 '퇴화형식(退化型式)'은 내포된 정치적 권위가 사라진 단계입니다. 2) 및 서봉총(瑞鳳塚)의 금관과 더불어 표지의 마디가 3단인 점이 특징입니다. 이 전단계에 속하며, 엊가지의 마디가 4단인 3단인 것에서 4단인 것으로 바뀌었고, 전합니다. 마지막인 '퇴화형식(退化型式)'은 전통제의 중심이었던 신궁(神宮)을 대신하여 토교를 공연하였고 더불어 국왕의 세속적 권위를 활동히 강화시켰습니다. 이로써 신라의 왕에게는 전통제의와 관련된 황금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지나고 있던 제의적 상징은 그대로 민간에 전승되어 정치적 권위가 사라진 채 무당의 신물로 남게 되었다고 볼 수 있습니다.

하나의 주제로 잘랐을 때
1,700 톤



데이터 구조화 방식

문제점 : 소제목 <h5>가 없거나 하나의 소단원이 너무 긴 경우 발생

큐레이터 추천 소장품

상장 검색

정보·보물 감색

3D 데이터 검색

수어 등장명

음성 안내

황남대총은 1973년부터 1975년까지 문화재관리국 문화재연구소(현재의 문화재청 국립문화재연구소)에서 발굴되었습니다. 신라의 둑무(甕鼻)무덤(甕鼻墓)으로 규모가 커서 그에 맞게 속하여, 귀금속 장신구와 허리띠 수입품 그리고 엉성한 철제풀과 금·Silver로 부장된 점에서 학자들은 이 무덤을 왕릉으로 판단하고 있습니다. 하지만 어느 왕의 능인지 정확히 알 수 없었지만, 1976년에 문화재위원회에 제출된 보고서에는 이 무덤을 '경주지 황남대총에서는 큰 무덤'이라 이름을 붙였습니다. 황남대총은 남북으로 두 개의 무덤을 뒤에 쌓았습니다. 먼저 만든 남쪽 무덤(南墓)은 왕의 능이었고, 북쪽 무덤(北墓)은 약간의 사신과 두고 나중에 잇낸 왕의 능이었기 때문에 밭굴을 부착되었습니다. 남쪽과 북쪽의 두 무덤은 둘다 지나친 남북무덤이란 점에서 같았습니다. 내부구조에서 남쪽 무덤은 시신을 모신 주파(主塚)와 더불어 부장품을 가득 차운 부락(副葬)을 따른 듯 점점해서 차이를 보였습니다. 북쪽은 왕비의 능이었다는 결론적 증거는 옛날 허리띠 구미개에 '夫人帶'라는 글씨가 새겨져 있었기 때문인데, 이 '부인'이 표기된 당시에 왕비를 끌고온 것입니다. 즉夫人은 왕비의 허리띠라는 의미를 지나고 할 수 있습니다. 황남대총은 북쪽에 안장된 왕비는 금관과 금으로 만든 꾸미 허리띠 이외에도 금발찌, 금반지, 금목걸이, 가죽꾸미개 등으로 장장하였습니다. 비록 비단 대포(大縷) 장구에 몇몇 직물을 기어 모두 썩어 없었으나, 남아있는 귀한 신진구는 생전에 노렸던 가장 화려한 복장을 입혀 장례가 이루어졌음을 짐작할 수 있습니다.

왕비의 금관은 신라에서 유래한 나뭇가지모양 금관의 전형을 보여 주고 있습니다. 세 개의 맞가지(對生枝)와 두 개의 가지가(苞生枝)로 조합된 세울장식(立葉)은 좁고 긴 머리띠에 놓이 솟아오도록 부착되었습니다. 맞가지는 경면과 좌우 측면에 세웠고, 앗가지는 위쪽으로 틀리스를 하도록 주위의 천지우에 축늘었습니다. 표면에는 무늬를 살피고 금은 보록과 달개비를 출마한 화려하게 수몄습니다. 비단과 고도 부르는 경우로 만든 곱은 목제 금관을 통을 개로 오른 달개비 하나하나씩 금실로 채웠습니다. 머리띠 아래에는 금은 고개에 사슬로 달개비를 출마한 화려하게 수몄습니다. 머리띠는 유기질로 된 으끈한 금관을 물었던 듯 합니다. 금은 고려의 금제 드리기는 착용한 사람의 양상을 염두해 두었음을 상기하고 있습니다. 이에 비해 대체로 남성의 거기와 드리기는 가느다리인 경우가 많습니다. 이 금관은 신라의 나뭇가지 모양 금관에서 맞가지와 앗가지가 조합된 최초의 것입니다.

황남대총에서 가장 의아한 수수께끼인 왕과 왕비가 지녔던 관(冠)을 들려는 문제입니다. 일반적인 생각과 전혀 다르게, 왕은 금동관을 지냈고 왕비는 금관을 지니고 있습니다. 구리관에 금을 한 금동관과 순수한 금관은 제작에서 엄청난 차이를 보입니다. 비단 왕의 관은 유연하게 곱은 옻을 닦은 금동관과는 별개로 불과합니다. 천지일 금동관에서 금관에 비해 결코 우회하지 않고 볼 수 없습니다. 뿐만 아니라 왕의 금관은 세울장식이 세 개의 가지를 끝입니다. 이러한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다.

신라의 '나뭇가지모양 관'은 세울 장식의 모양에 따라 변화를 3단계로 나눌 수 있습니다. 처음에는 머리띠에 세 개의 마디(마디)를 네워서 나온 마디가 '나뭇가지모양(나뭇가지모양)'입니다. 그 다음에는 세 개의 마디를 구리판을 적절히 잘라 만든 동관(銅冠)입니다. 동관의 착용처는 함께 부장된 방울 등을 둘러 끌 때 이전 단계의 정치적 권위가 사라진 무당일 가능성 있습니다.

이러한 변화상은 '나뭇가지모양 관'이 지니고 있는 상징성을 짐작할 수 있게 합니다. 신라의 왕과 왕족은 기록에 따르면 한때 제사장이었으며, 마립간식기(麻立干食期)까지만 해도 국회의 제사를 주관하였습니다. 황금으로 만든 '나뭇가지모양 관'은 신성한 나무를 상징하는 듯합니다. 이는 흔치 않은 국가의 제사를 주관하는 신라의 마립간과 그 일족의 성격에 부합하는 듯합니다. 이처럼 신라의 왕에게는 전통제국과 관련된 향금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지니고 있던 제적의 성정은 그대로 민간에 전溯되어 정치적 권위가 사라진 채 무당의 물품으로 남게 되었습니다. 볼 수 있습니다.

왕비의 금관. 경주 황남대총 복원 출토, 높이 27.3cm, 낙보

| 짐작할 수 있게 합니다.

나눌 수 있습니다. 처음에는 머리띠에 세 가지가지가 밖으로 벌어진 것에서 점자 카랑 헤이타이으로 '시원형식(始原形式)'으로 분류되는 남분에서 왕이 사용한 금동관과 함께 어울리게 험식과도 이야기됩니다. '나뭇 것인데, 가장 다른 점은 험태이므로 '표준 험(標准形)'입니다. 머리띠에 세 가지가지는 위로 높게 더해졌습니다. 측면에서 나무와 사슴뿔을 하늘과 향을 '걸과' 모두 직각으로 꺾인 형태입니다. | 서·봉서(瑞鳳瑞)는 금관과 더불어 표지의 마지막 3단인 점이 특징입니다. '나뭇 것과 함께 하늘과 땅, 머리띠가 미だ리가 4단인 3단인 것에서 4단인 것으로 바뀌었고, 전니다. 미다리인 '화학형식(花化形式)'이라고도 합니다. | 내포된 정치적 권위가 사라진 단계이므로

제작되는 험식 형식입니다. 황금으로 만든 '나뭇가지모양 관'은 신성한 나무를 상징하는 듯합니다. 마립간식기(麻立干食期)로 대신하여 구왕의 세속적 권위를 활동하고 강화되었습니다. 이로써 신라의 왕에게는 전통제국과 관련된 향금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지니고 있던 제적의 성정은 그대로 민간에 전溯되어 정치적 권위가 사라진 채 무당의 물품으로 남게 되었습니다. 볼 수 있습니다.

하나의 주제로 잘랐을 때 1,700 > 512 (BERT)

데이터 구조화 방식

문제점 : 소제목 <h5>가 없거나 하나의 소단원이 너무 긴 경우 발생

큐레이터 추천 소장품 소장품 검색 국보·보물 검색 3D 데이터 검색 수어 동영상 음성 안내

황남대총은 1973년부터 1975년까지 문화재관리국 문화재연구소(현재의 문화재청 국립문화재연구소)에서 발굴하였습니다. 신라의 돌무지덧널무덤 가운데 가장 큰 것은 속하여, 귀금속 장신구와 희귀한 수입품 그리고 엄청난 양의 철제품과 길그릇이 부장된 점에서 학자들은 이 무덤을 왕릉으로 판단하고 있습니다. 하지만 어느 왕의 능인지 정확히 알 수 없었으므로, 1976년에 문화재위원회는 이 무덤을 황남동에 있는 큰 무덤이란 이름으로 '황남대총'이라는 이름을 붙였습니다. 황남대총은 남쪽으로 두 개의 무덤을 뒷부분에 놓았습니다. 그래서 만든 남쪽 무덤(南墳)은 왕의 능이었고, 북쪽 무덤(北墳)은 약간의 차이를 두고 나중에 보면 왕의 능이었으나, 남부구조에서 남쪽 무덤은 통해 밝혀졌습니다. 남쪽과 북쪽의 무덤 모양은 돌무지덧널무덤이란 점에서 같으나, 남부구조에서 남쪽 무덤은 시선을 모신 주과(主佛)과 더불어 부장품을 가진 채로 부과(副佛)를 떠로 둔 점에서 차이를 보았습니다. 북쪽이 왕비의 능이었는지 결정적 증거는 온전히 꾸미개가(夫人帶)라는 글씨가 새겨져 있었기 때문입니다. 이 외인이라면 표현이 당시 왕비를 말하는 것입니다. 즉 '夫人帶'은 왕비의 허리띠라는 의미를 지닌다고 할 수 있습니다. 황남대총 북분에 안장된 왕비는 금관과 금모로 끌금짜, 금반지, 금목걸이, 기술꾸미개 등으로 치장하였습니다. 비록 비단포와 장신구에 대비해 직물이 거의 모두 써되어졌으나, 남아있는 귀한 장신구는 생전에 누웠던 가장 화려한 복장을 입혀 장례가 이루어졌음을 짐작할 수 있습니다.

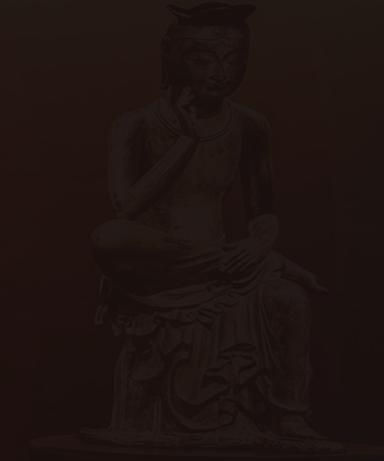
왕비의 금관은 신라에서 유행한 나뭇가지모양 금관의 전형을 보여 주고 있습니다. 세 개의 맞가지[臘生枝]와 두 개의 엇가지[互生枝]로 조합된 세움장식[立輪]을 놓고 긴 머리띠에 높이 솟아오르도록 부착되었습니다. 맞가지는 정면과 좌우측에 세워졌고, 엇가지는 위쪽으로 비스듬하도록 후면과 좌우측에 붙였습니다. 표면에는 무늬를 새기고 곰은목과 달개를 충실히 새겼습니다. 뒷면에는 무늬를 새기고 곰은목과 달개를 충실히 새겼습니다. 비취라고도 부른 경목으로 '앙다'라 응어리로 금봉은은[세늘경식]이 새겨 맞가지 분입니다. 이러한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다.

금은 고리에 사슬로 달개를 얹은 한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다.

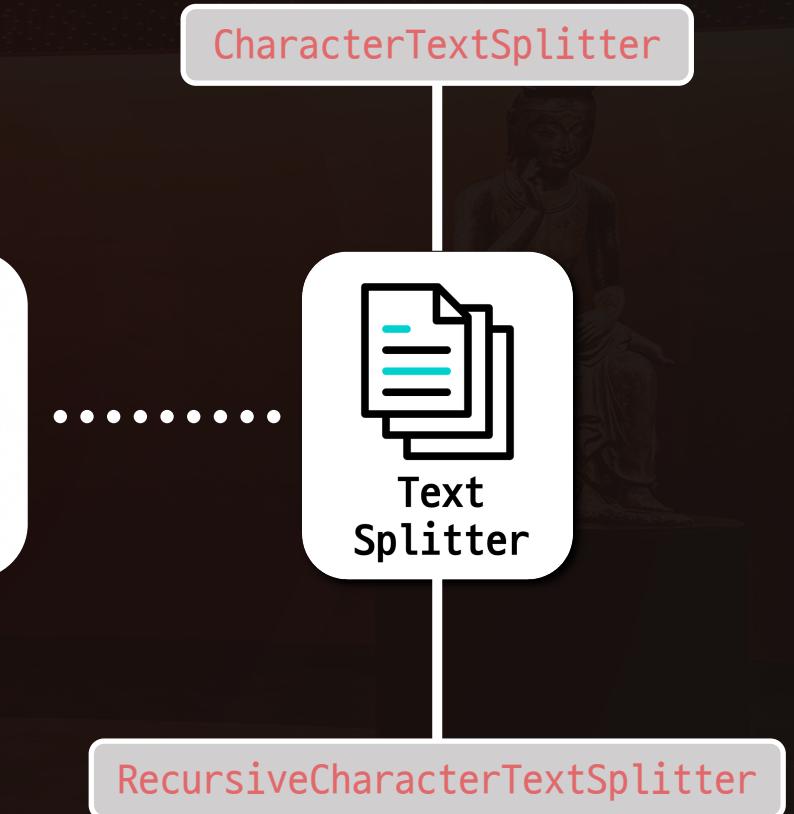
신라의 '나뭇가지모양 관'은 세울 장식의 모양에 따라 변화를 3단계로 나눌 수 있습니다. 처음에는 머리띠에 세 개의 맞가지만을 세웠습니다. 맞가지의 모양은 자연스런 나뭇가지처럼 가지가 밖으로 벌어진 것에서 점차 직각으로 꺾이는 것으로 변화합니다. 자연의 나무를 형상화한 듯하며, 초기 형태이므로 '시원형식(始原型式)'으로 분류하는데, 경주 교동사에서 발견된 금관이 대표적인입니다. 이 형식의 마지막은 황남대총 남분에서 왕이 착용한 금동관입니다. 세워 장식의 세 개의 맞가지 조합된 것인데, 가장 다듬어진 형태이므로 '표준형식(標準型式)'으로 분류합니다. 세워 장식의 세 개의 엇가지 조합된 것인데, 가장 뒤늦게 더해졌으므로 이를 굳이 사슴뿔을 형상화했다고 볼 여지가 적습니다. 다만 상징 측면에서 나무와 사슴뿔은 하늘과 땅을 이어주는 맥락의 좋은 의미를 지닙니다. 표준형식의 이후의 맞가지 조합된 것에서는 한결같이 모두 직각으로 꺾여진 형태입니다. 표준형식은 황남대총 북분 금관에서 처음으로 완성되며, 금관총(金冠塔) 및 서봉총(瑞鳳塚)의 금관과 더불어 표준형식이 성립되는 단계에 속합니다. 표준형식은 성립단계의 금관은 맞가지의 마디가 3단인 점이 특징입니다. 이후에 만들어진 천마총(天馬塚) 및 금령총(金鶻塚)은 금관과 표준형식을 혼전단계에 속하며, 맞가지의 마디가 4단인 특징을 지닙니다. 즉 '나뭇가지모양 관'의 표준형식은 맞가지의 마디가 3단인 것에서 4단인 것으로 바뀌었고, 전체적으로 점점 뾰족해지며 화려해지는 모습으로 변형되는 일 수 있습니다. 마지막인 '퇴화형식(退化型式)'은 축축해진데 더 이상 금관이 만만하지거나 오히려 급격하게 간단해지는 방향으로 변화합니다. 퇴화형식에서도 가장 마지막은 구리관을 적당히 잘라 만든 동관(銅冠)입니다. 동관의 학용자는 함께 부장된 방울 등으로 볼 때 이전 단계의 정치적 권위가 사라진 무당일 가능성이 있습니다.

이러한 변화상은 '나뭇가지모양 관'이 지니고 있는 상징성을 짐작할 수 있게 합니다. 신라의 왕과 왕족은 기록에 따르면 해당성이 있으며, 마립간시기(麻立干時期)까지 꾸준히 국가의 제의를 주관하였습니다. 황금으로 만든 '나뭇가지모양 관'은 신성한 나무를 상징하는 듯합니다. 이는 통치자이자 국가의 제의를 주관하는 신라의 마립간과 그 일족의 성격에 부합하는 도안이라고 해석됩니다. 마립간시기 직후인 중고기의 법왕왕(法興王)은 전통제의의 중심이었던 신궁(神宮)을 대신하여 불교를 공인하였고 더불어 국왕의 세속적 권위를 높임에 강화시켰습니다. 이로써 신라의 왕에게는 전통제의와 관련된 황금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지니고 있던 제의적 성질은 그대로 민간에 전승되어 정치적 권위가 사라진 채 무당의 신물로 남게 되었다고 볼 수 있습니다.

} 문단으로 잘라도
590 토큰 > 512



Solution



데이터 구조화 방식

아니라 왕의 금동관은 제품상적이 제 개의 맞가지 문입니다. 이러한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다.

신라의 '나뭇가지모양 관'은 세움 장식의 모양에 따라 변화를 3단계로 나눌 수 있습니다. 처음에는 머리띠에 세 개의 맞가지만을 세웠습니다. 맞가지의 모양은 자연스런 나뭇가지처럼 가지가 밖으로 벌어진 것에서 점차 직각으로 꺾이는 것으로 변화합니다. 자연의 나무를 형상화한 듯하며, 초기 형태이므로 '시원형식(始原型式)'으로 분류하는데, 경주 교동에서 발견된 금관이 대표적입니다. 이 형식의 마지막은 황남대총 남분에서 왕이 착용한 금동관입니다. 세움 장식은 맞가지 뿐이지만, 그 형태가 직각에 꺾여 있어 뒤이은 형식과도 이어집니다. '나뭇가지모양 관'의 전형은 세 개의 맞가지에 더해 두 개의 엇가지를 조합한 것인데, 가장 다듬어진 형태이므로 '표준형식(標準型式)'으로 분류합니다. 새로 조합된 엇가지는 사슴뿔을 본떴다고도 합니다. 엇가지는 뒤늦게 더해졌으므로 이를 굳이 사슴뿔을 형상화했다고 볼 여지가 적습니다. 다만 상징 측면에서 나무와 사슴뿔은 하늘과 땅을 이어주는 매개라는 같은 의미를 지닙니다. 이 단계 이후의 맞가지는 한결같이 모두 직각으로 꺾인 형태입니다. 표준형식은 황남대총 북분 금관에서 처음으로 완성되며, 금관총(金冠塚) 및 서봉총(瑞鳳塚)의 금관과 더불어 표준형식이 성립하는 단계에 속합니다. 표준형식 성립단계의 금관은 맞가지의 마디가 3단인 점이 특징입니다. 이후에 만들어진 천마총(天馬塚) 및 금령총(金鶯塚)의 금관은 표준형식 발전단계에 속하며, 맞가지의 마디가 4단인 특징을 지닙니다. 즉 '나뭇가지모양 관'의 표준형식은 맞가지의 마디가 3단인 것에서 4단인 것으로 바뀌었고, 전처적으로 점점 빠져지고 화려해지는 모습으로 변하였음을 알 수 있습니다. 마지막인 '퇴화형식(退化型式)'은 추측하건데 더 이상 금관이 만들어지지 않는 단계로 볼 수 있습니다. 관에 내포된 정치적 권위가 사라진 단계이므로 필요 이상으로 꾸밈이 많아지거나 오히려 급격하게 간단해지는 방향으로 변화합니다. 퇴화형식에서도 가장 마지막은 구리판을 적당히 잘라 만든 동관(銅冠)입니다. 동관의 착용자는 함께 부장된 방울 등으로 볼 때 이전 단계의 정치적 권위가 사라진 무당일 가능성성이 있습니다.

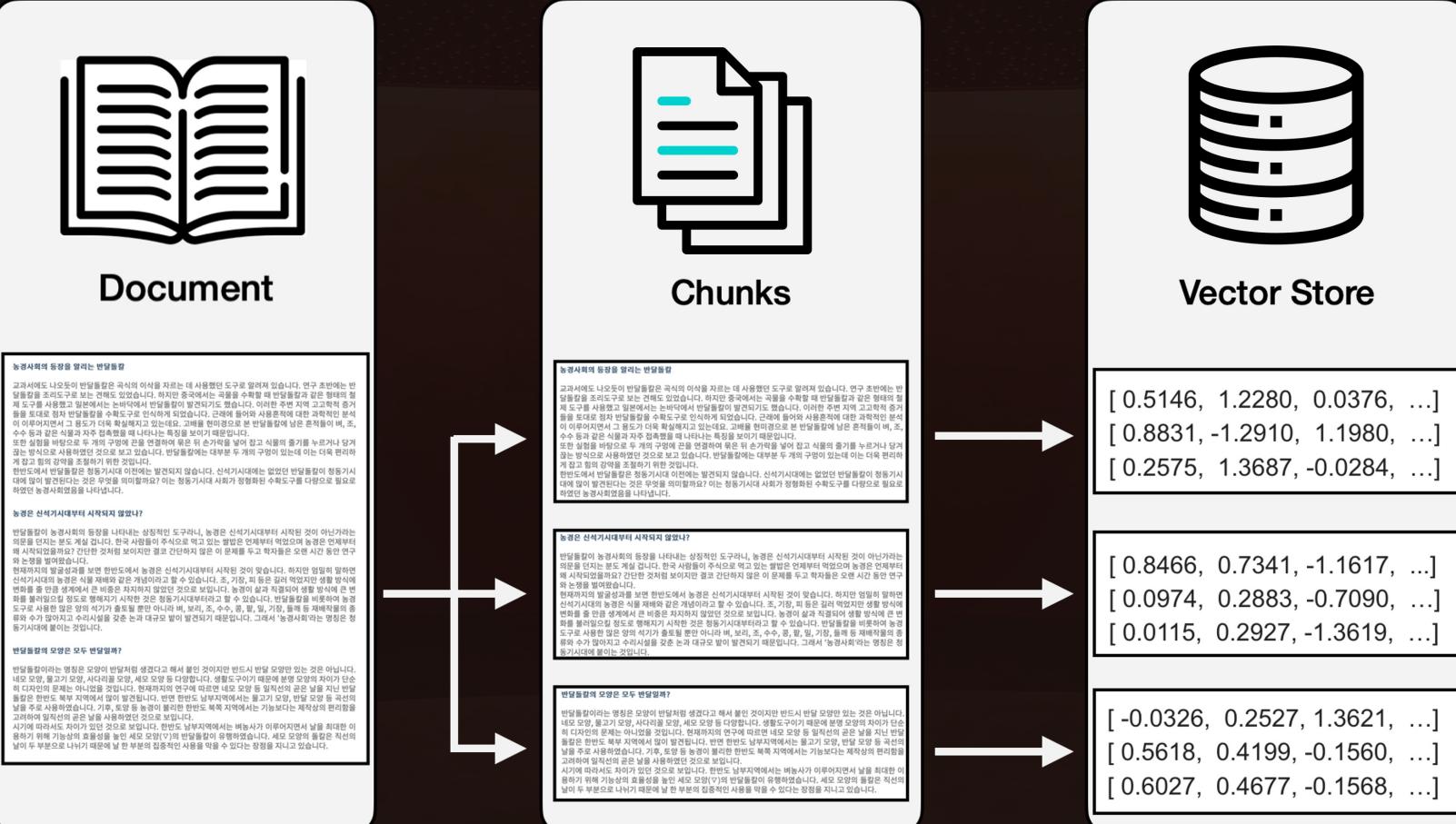
이러한 변화상은 '나뭇가지모양 관'이 지니고 있는 상징성을 짐작할 수 있게 합니다. 신라의 왕과 왕족은 기록에 따르면 한때 제사장이었으며, 마립간시기(麻立干時期)까지 꾸준히 국가의 제의를 주관하였습니다. 황금으로 만든 '나뭇가지모양 관'은 신성한 나무를 상징하는 듯합니다. 이는 통치자이자 국가의 제의를 주관하는 신라의 마립간과 그 일족의 성격에 부합하는 도안이라고 해석됩니다. 마립간시기 직후인 중고기의 법흥왕(法興王)은 전통제의의 중심이었던 신궁(神宮)을 대신하여 불교를 공인하였고 더불어 국왕의 세속적 권위를 월등히 강화시켰습니다. 이로써 신라의 왕에게는 전통제의와 관련된 황금으로 만든 '나뭇가지모양 관'이 더 이상 필요 없게 된 듯합니다. 다만 나뭇가지모양 관이 지니고 있던 제의적 상징은 그대로 민간에 전승되어 정치적 권위가 사라진 채 무당의 신물로 남게 되었다고 볼 수 있습니다.

{"description": "신라의 '나뭇가지모양 관'은 세움 장식의 모양에 따라 변화를 3단계로 나눌 수 있습니다. 처음에는 머리띠에 세 개의 맞가지만을 세웠습니다. 이러한 왕과 왕비의 관은 신라의 '나뭇가지모양 관'이 어떻게 변화하였는지를 짐작할 수 있게 합니다."}

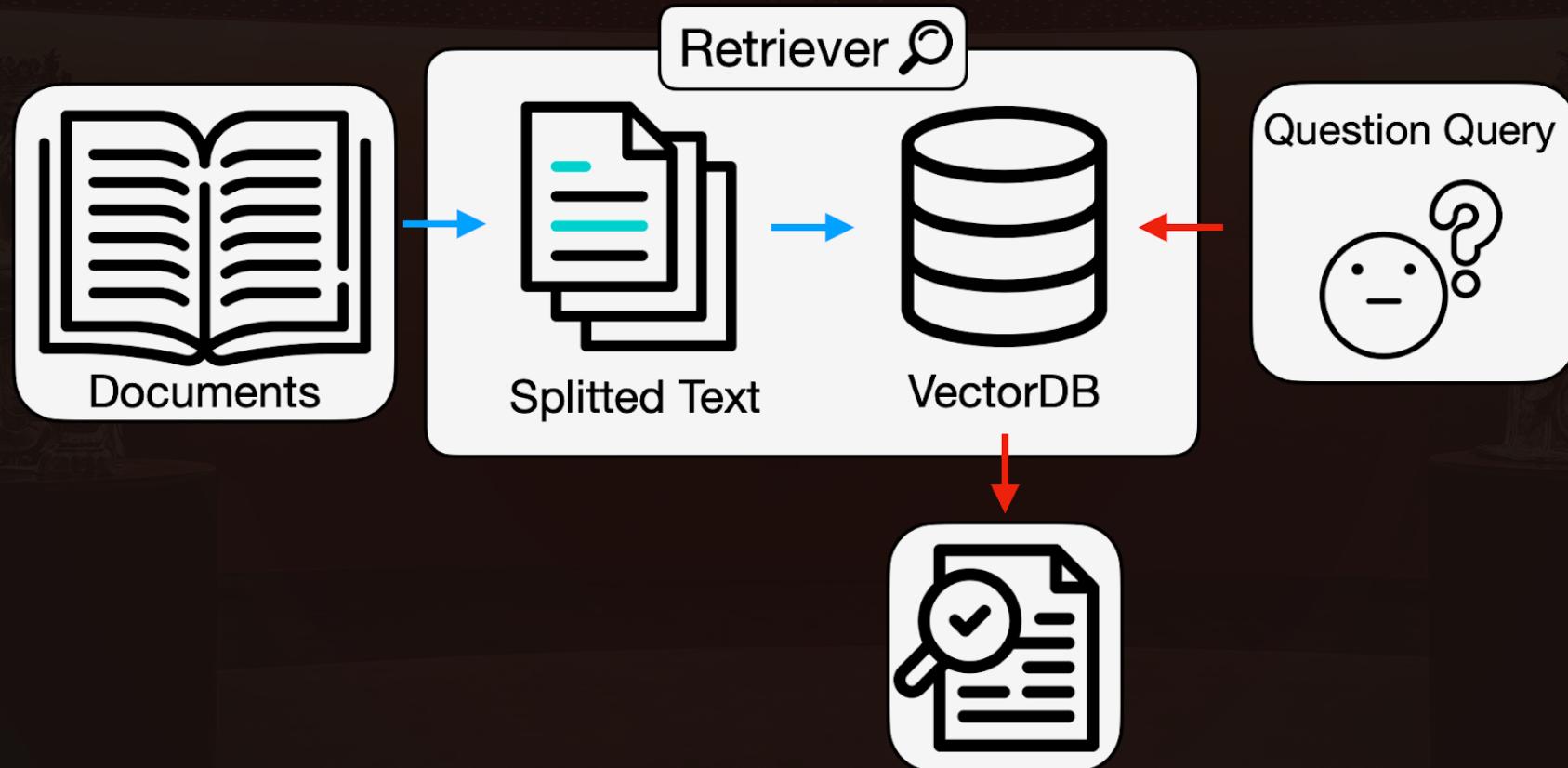
{"description": "다만 상징 측면에서 나무와 사슴뿔은 하늘과 땅을 이어주는 매개라는 같은 의미를 지닙니다. 이 단계 이후의 맞가지는 (후략) 권위가 사라진 무당일 가능성성이 있습니다."}

```
splitter = CharacterTextSplitter.from_huggingface_tokenizer(chunk_size=380, chunk_overlap=50,
tokenizer=tokenizer, separators=["한편", "이 밖에도", "다만", ". ", ..(생략)], keep_separator=True)
```

데이터 벡터화 및 저장



Retriever의 작동 구조



전통적인 Retrieval 방식

Term-matching method (Sparse Embedding)

TF-IDF

Term Frequency - Inverse Document Frequency

BM25

Best Matching 25

BM25를 활용한 RAG 모델 성능

User Query : “금으로 만들어진 유물 중 하나를 알려주세요.”

유사도를 기준으로 top 5의 passage들을 찾아오게 함

BM25를 활용한 RAG 모델 성능

User Query : "금으로 만들어진 유물 중 하나를 알려주세요."

```
bm25.get_top_n(query = "금으로 만들어진 유물 중 하나를 알려주세요.")
```

['북청 여진문자 석각(北青女眞文字石刻) 탁본-함경도에서 발견된 금나라 마애명(磨崖銘)을 찾아갑니다. 그리곤 사람이 도무지 살 수 없을 것 같은, 산양이나 뛰놀 것 같은 바위를 찾았습니다. 그가 사람의 손길로 쪼듯한 무늬가 보입니다. 무늬? 아니 무늬가 아니라 글자입니다. 글자? 아니 글자도 아니고, 한글은 더더욱 아니고요. 사람들은 이윽고 송방망이를 만들어 물을 적시고, 종이를 두드렸습니다. 그리고는 다른 소반마이를 두 개 만들었습니다. 하나에 먹물을 등쪽 묻히더니 다른 하나를 꾹꾹 문지릅니다. 그렇게 문질러 먹종이 위를 탕탕 두들깁니다. 금세 글자의 윤곽이 또렷해집니다. 이성계 밭원 사리갖춤#n1932년 10월 6일, 금강산 월출봉(1580m)에서 발견된 금나라 마애명(磨崖銘)이네.']

['이성계 밭원 사리갖춤#n1932년 10월 6일, 금강산 월출봉(1580m)에서 발견된 금나라 마애명(磨崖銘)을 상자 하나를 발견하였습니다. 그 안에는 사리갖춤이 들어 있었는데, 이성계와 부인 강원 이들을 비롯한 1만여 명이 미륵을 기다리며 금강산 아도봉(adobong)에 서니꽃봉을 조망하는 대봉의 정운이 새겨져 있었습니다. 당초 비로봉에 놓았던 사리갖춤이 어떻게 월출봉에서 발견됐는지는 '묵소거사자찬(默笑居士自讚), 김정희#n "나와 이재와 추사는 사람들이 말하는 석교(石交 금석처럼 두텁고 견고한 유품)를 만난다면 정치적 득실과 인물의 시비에 대해서는 말하지 않지 않는다. 다만 고금에 대해 이야기하고 서화를 풍평할 뿐이다. 하루라도 보지 않으면 문득 슬퍼하(石交 금석처럼 두텁고 견고한 성명과 자호가 모두 그곳에 있으니 마치 그 사람을 볼 하나님을 구하면 오른쪽 왼쪽 예백에 모두 두 사람의 도장을 찍어 얼굴을 대신하는 자료로 여겼다."]

['활빈당 발령-낡은 종이에 낡은 민초(民草)의 목소리#n1903년 겨울, 해가 어느덧 산 너머로 넘어가는 밤 충청북도 회인(懷仁, 지금의 보은 일대)에 살던 부자(富者) 정인원은 자기 집 사랑채 안 그저 평안히 있는 것처럼 보이지만, 가까이서 보니 종잇장 하나를 하나를 들고 벌벌 떨고 있다. 그는 그의 어깨너머로 그 글을 읽어봅시다.']

['빗살무늬토기#n신석기시대에 만들어진 토기는 흔히 빗살무늬토기로 불려서 썼습니다. 이시간 구석기시대에서 가장 먼저 만들어진 토기는 무늬가 없는 '고산리식 토기' 라 불리는 것입니다. 제주도 기를 빚을 때 풀과 같은 유기물을 첨가하여 만들었는데, 구석기시대 석기제작기술을 보여주는 돌 화살촉과 함께 확인되며 기원전 8,000년경으로 추정되고 있습니다.']

| 대에 만들어진 토기는 흔히 빗살무늬 토기로 불리워 썼습니다.

BM25를 활용한 RAG 모델 성능

User Query : “외국인의 작품이 있을까?”

```
bm25.get_top_n(query = "외국인의 작품이 있을까?")
```

[‘동래부사집왜사도’
n동래부에서 일본 사신의 방문, 특히 정기적으로 치러졌던 연례 충사는 고율의 위상을 규정짓는 특별하고도 일상화된 이벤트였을 것입니다. 행사는 주요 등장인물이 ‘외국인’이라는 점은 더욱 흥미롭습니다.
외국인의 모습이나 이국의 풍물습속을 소재로 하는 그림은 중국의 오래된 회화 관련 서적에도 보이는 풍속화의 고전적 주제라고 할 수 있습니다. 북송 선화 연간에 고려에 다녀간 중국 사신 서공(徐公)은 외국에 대한 정보수집의 차원에서 고려의 문물을 사각화한 『선화봉사고려도경(宣和奉使高麗圖經)』을 지었습니다. 중국에 온 사신들의 모습을 그린 <직공도(職貢圖)>는 주변 이민족에 대한 지배력을 확고히 하기 위해 제작한 정치적 의도의 풍속화라고도 할 수 있지만, 소재적인 측면에서 보면 이국적인 것, 미지의 것에 대한 인간의 호기심이 기저에 있음을 무시할 수는 없을 것입니다’.

. 행사의 주요 등장인물이 ‘외국인’이라는 점은 더욱 흥미롭습니다.

Dense Passage Retrieval (DPR)

.CL] 30 Sep 2020

<https://arxiv.org/pdf/2004.04906.pdf>

Facebook AI (Meta), Submitted on 10 Apr 2020 (v1), last revised 30 Sep 2020 (this version, v3)

Dense Passage Retrieval for Open-Domain Question Answering

Vladimir Karpukhin*, Barlas Oğuz*, Sewon Min†, Patrick Lewis,
Ledell Wu, Sergey Edunov, Danqi Chen‡, Wen-tau Yih

Facebook AI †University of Washington ‡Princeton University
{vladk, barlaso, plewis, ledell, edunov, scottyih}@fb.com
sewon@cs.washington.edu
danqic@cs.princeton.edu

Abstract

Open-domain question answering relies on efficient passage retrieval to select candidate contexts, where traditional sparse vector space models, such as TF-IDF or BM25, are the de facto method. In this work, we show that retrieval can be practically implemented using *dense* representations alone, where embeddings are learned from a small number of questions and passages by a simple dual-encoder framework. When evaluated on a wide range of open-domain QA datasets, our dense retriever outperforms a strong Lucene-BM25 system greatly by 9%-19% absolute in terms of top-20 passage retrieval accuracy and

Retrieval in open-domain QA is usually implemented using TF-IDF or BM25 (Robertson and Zaragoza, 2009), which matches keywords efficiently with an inverted index and can be seen as representing the question and context in high-dimensional, sparse vectors (with weighting). Conversely, the *dense*, latent semantic encoding is *complementary* to sparse representations by design. For example, synonyms or paraphrases that consist of completely different tokens may still be mapped to vectors close to each other. Consider the question “Who is the bad guy in lord of the rings?”, which can be answered from the context “Sala Baker is best known for portraying the villain Sauron in the Lord

Dense Passage Retrieval (DPR)

Positive Passage Negative Passage

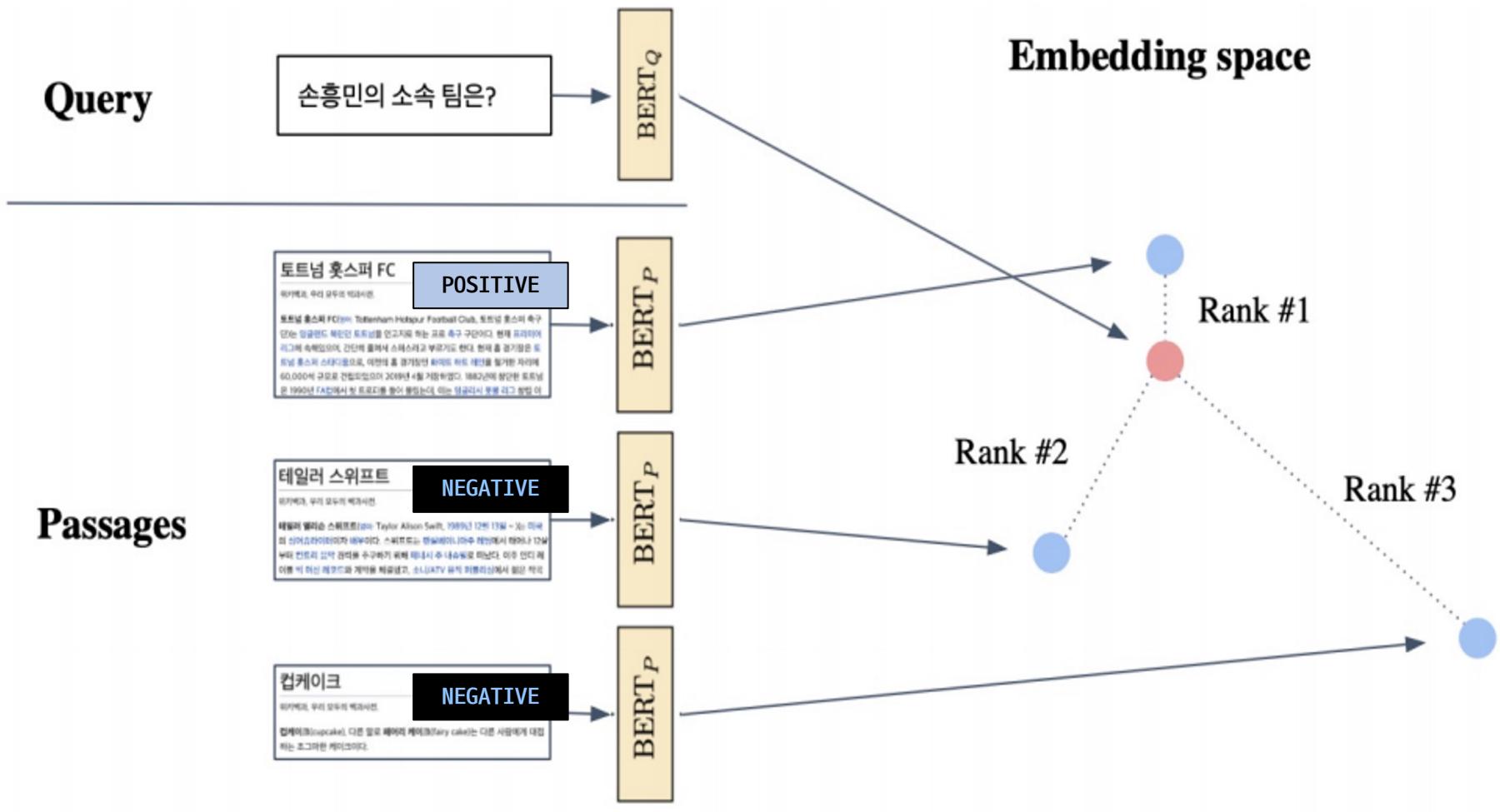
$$D = \{< q_i, p_i^+, p_{i,1}^-, \dots, p_{i,n}^- >\}_{i=1}^m$$

$$L(q_i, p_i^+, p_{i,1}^-, \dots, p_{i,n}^-) = -\log \frac{e^{sim(q_i, p_i^+)}}{e^{sim(q_i, p_i^+)} + \sum_{j=1}^n e^{sim(q_i, p_{i,j}^-)}}$$

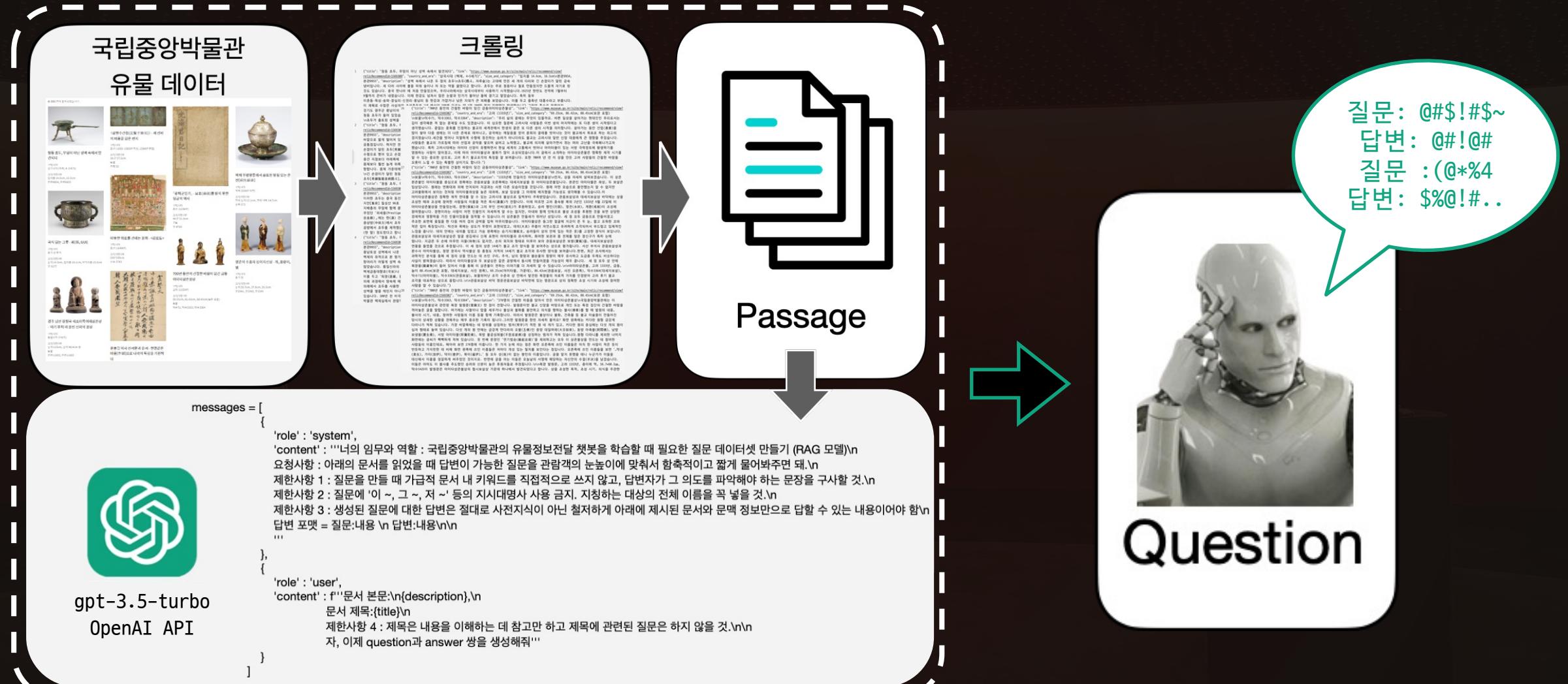
using the dot product

$$sim(q, p) = E_Q(q)^T E_P(p)$$

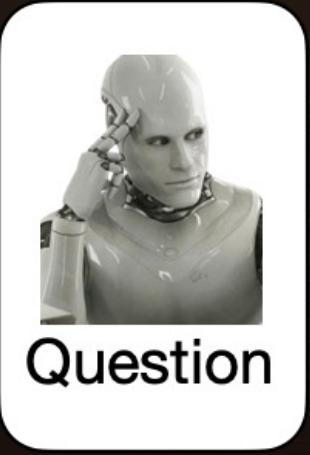
DPR 학습 방식 예시 이미지



DPR 학습을 위한 데이터셋 생성



DPR 학습을 위한 데이터셋 생성



BM25

Hard Negative Passage 생성

데이터 구조화

```

def make_dataset(self, num_data):
    question_data = []
    pos_passage_data = []
    negative_passage_data = []
    negative_question_data = []
    for row, questions in tqdm(enumerate(self.questions)):
        for col, question in tqdm(enumerate(questions)):
            top_idx = self._get_negative_passage_bm25(row, col)
            if top_idx == -1:
                continue
            else:
                if len(self.questions[top_idx]) == 0:
                    continue
                if len(self.questions[top_idx]) == 1:
                    negative_question_data.append(self.questions[top_idx][0])
                else:
                    negative_question_data.append(self.questions[top_idx][random.randint(0, len(self.questions[top_idx]) - 1)])
            question_data.append(question)
            pos_passage_data.append(self.contexts[row])
            negative_passage_data.append(self.contexts[top_idx])
    if len(question_data) >= num_data:
        break

#shuffle
index_list = list(range(len(question_data)))
random.shuffle(index_list)

questions = [question_data[idx] for idx in index_list]
pos_passages = [pos_passage_data[idx] for idx in index_list]
negative_passages = [negative_passage_data[idx] for idx in index_list]

return Dataset.from_pandas(pd.DataFrame({'pos_question': questions[:num_data],
                                         'pos_passage': pos_passages[:num_data],
                                         'neg_passage': negative_passages[:num_data]}))

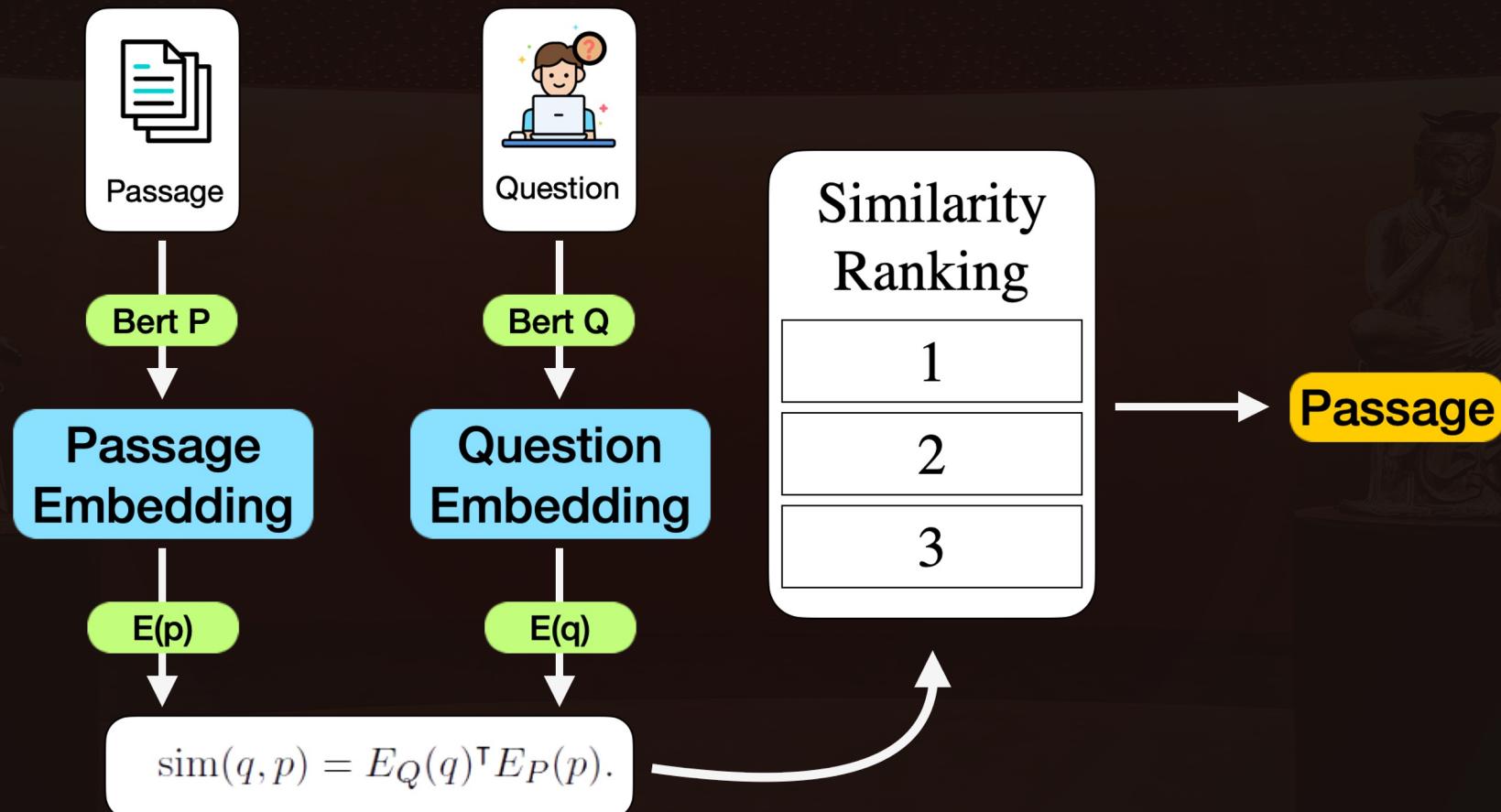
```

DPR 데이터셋 구조

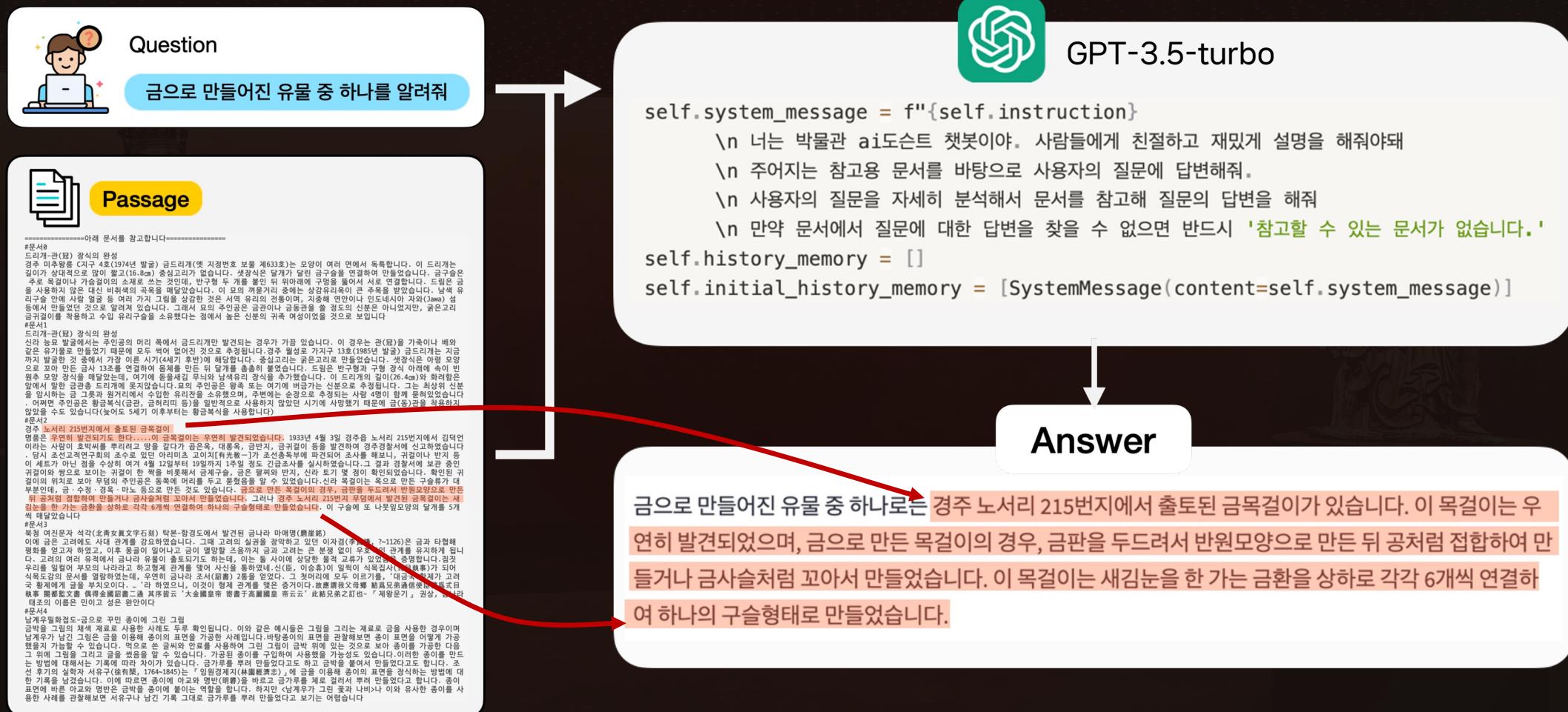
```
"features": {  
    "question": {  
        "dtype": "string",  
        "_type": "Value"  
    },  
    "pos_passage": {  
        "dtype": "string",  
        "_type": "Value"  
    },  
    "hard_neg_passage": {  
        "dtype": "string",  
        "_type": "Value"  
    }  
}
```

Dataset	Train	Test
AI hub	99,898	2,697
Museum	95,000	2,697
AI hub + Museum	194,898	2,697

Retriever 작동 방식 - Similarity



Retriever 작동 방식 - LLM 연결



Streamlit으로 구현한 챗봇 프로토타입

국립중앙박물관 챗봇 ★

안녕하세요! 저는 박물관 AI도슨트 슈팅스타입니다.

박물관에 처음 왔는데 추천해줄만한 유물이 있을까?

박물관에 처음 방문하신다면, 구석기시대의 유물을 보는 것을 추천드립니다. 구석기시대 최첨단 도구인 작은 돌날이나 청자 구름 학 무늬 매병과 같은 유물들이 흥미롭고 가치 있는 전시물일 것입니다. 이러한 유물들은 고대 문화와 인류의 역사를 엿볼 수 있는 소중한 자료들입니다. 따라서 구석기실과 특별전 전시실을 방문해 보시는 것을 권해드립니다.

구석기 시대의 유물을 소개해줘

구석기 시대의 유물은 주로 돌로 만들어진 도구들이 많이 발견되었습니다. 이러한 돌 도구들은 주로 사냥이나 식량을 얻기 위해 사용되었으며, 주로 돌날이나 주먹도끼 등이 있습니다. 또한, 신석기시대에는 해양 자원을 활용하는 모습도 보여 고래 뼈나 조개껍질 등이 발견되었습니다. 이러한 유물들은 당시 사람들의 생활 방식과 문화를 엿볼 수 있는 중요한 자료입니다.

주먹도끼에 대해서 알려줄래?

Your message ➤



Streamlit으로 구현한 챗봇 프로토타입

국립중앙박물관 챗봇

안녕하세요! 저는 박물관 AI도슨트 슈팅스타입니다.

박물관에 처음왔는데 추천해줄만한 유물이 있을까?

박물관에 처음 방문하신다면, 구석기시대의 유물을 보는 것을 추천드립니다. 구석기시대 최첨단 도구인 작은 돌날이나 청자 구름 학 무늬 매병과 같은 유물들이 흥미롭고 가치 있는 전시물일 것입니다. 이러한 유물들은 고대 문화와 인류의 역사를 엿볼 수 있는 소중한 자료들입니다. 따라서 구석기실과 특별전 전시실을 방문해 보시는 것을 권해드립니다.

구석기 시대의 유물을 소개해줘

구석기 시대의 유물은 주로 돌로 만들어진 도구들이 많이 발견되었습니다. 이러한 돌 도구들은 주로 사냥이나 식량을 얻기 위해 사용되었으며, 주로 돌날이나 주먹도끼 등이 있습니다. 또한, 신석기시대에는 해양 자원을 활용하는 모습도 보여 고래 뼈나 조개껍질 등이 발견되었습니다. 이러한 유물들은 당시 사람들의 생활 방식과 문화를 엿볼 수 있는 중요한 자료입니다.

주먹도끼에 대해서 알려줄래?

Your message ▶

Query

```
주먹도끼에 대해서 알려줄래? [0.41070669889450073, 0.4084071218967438, 0.406744420528411  
87, 0.38531559705734253, 0.3840537965297699]  
threshold를 넘는 index : [0, 1, 2, 3, 4]
```

Passage Retrieval

=====아래 문서를 참고합니다=====

#문서0
연천 전곡리 주먹도끼
그렉 보웬(Greg L. Bowen)은 미국 캘리포니아 빅터 벌리 대학에서 고고학을 공부하다가 1974년 미군에 입대하면서 한국의 동두천에서 주한 미공군 하사관으로 근무하게 되었습니다. 그는 1978년 한탄강 유원지에 한국인 애인과 함께 산책을 갔다가 토기편을 발견하고 그 주위를 조사하면서 주먹도끼를 발견하게 되었습니다. 그때만 해도 이 주먹도끼가 동아시아 구석기 고고학에 있어 기념비적인 발견이 되리라고는 그 누구도 상상하지 못했습니다

#문서1
주먹도끼
주먹도끼의 제작자는 인류의 직계 조상인 곧선사람(Homo erectus)입니다. 곧선사람 이전까지의 인류 조상들은 아프리카에서 살았습니다. 그러나 지금으로부터 약 160만 년 전 쯤에 출현한 곧선사람들은 불을 능숙하게 다듬으로써 비로소 아프리카를 벗어나 유럽과 아시아로까지 삶의 영역을 확장시킬 수 있었습니다. 곧선사람의 화석 출토 범위와 주먹도끼가 발견되는 지역의 범위는 대체로 일치합니다

#문서2
주먹도끼
주먹도끼 -> 국적/시대:한국(韓國)-구석기(舊石器), 크기길이 23.6cm, 문화재구분정보없음, 소장품번호신수(新收) 019143 00000

#문서3
연천 전곡리 주먹도끼
주먹도끼는 형태적으로 끝이 뾰족하거나 전체적으로 등근 타원형의 석기를 말하며, 기능적으로는 주먹에 쥐고 사용했던 석기를 말합니다

#문서4
연천 전곡리 주먹도끼
서양의 주먹도끼는 아슬리안(Acheulian)으로 불리는데, 프랑스 북서부 솜드강 강변에 위치한 생 아슬(St. Acheul)에서 다양한의 석기가 확인되면서 불여진 명칭입니다. 유럽과 아프리카에서는 개별 도구를 나타내는 주먹도끼라는 용어보다는 전체적인 석기군의 특성을 아우르는 아슬리안이라는 용어를 많이 사용합니다. 서양에서 부르는 고전적인 아슬리안 주먹도끼는 아몬드나 타원의 평면 형태를 가지고 있으며 끝이 뾰족하며, 단면은 양쪽이 볼록한 렌즈 모양입니다. 석기 주변부를 따라서 날카로운 날이 있고, 양면 전체를 빼내어 만듭니다

Streamlit으로 구현한 챗봇 프로토타입

국립중앙박물관 챗봇 ★

안녕하세요! 저는 박물관 AI도슨트 슈팅스타입니다.

박물관에 처음왔는데 추천해줄만한 유물이 있을까?

박물관에 처음 방문하신다면, 구석기시대의 유물을 보는 것을 추천드립니다. 구석기시대 최첨단 도구인 작은 돌날이나 청자 구름 학 무늬 병과 같은 유물들이 흥미롭고 가치 있는 전시물일 것입니다. 이러한 유물들은 고대 문화와 인류의 역사를 엿볼 수 있는 소중한 자료들입니다. 따라서 구석기실과 특별전 전시실을 방문해보시는 것을 권해드립니다.

LLM의 Answer

주먹도끼는 골선사람(Homo erectus)이 만든 도구로, 주먹에 쥐고 사용하는 석기를 말합니다. 이 도구는 주로 돌로 만들어졌으며, 끝이 뾰족하거나 전체적으로 둥근 타원형을 가지고 있습니다. 주먹도끼는 골선사람의 불을 다루고 도구를 만들며 생활 영역을 확대하는 데 사용되었습니다. 서양에서는 아슬리안(Acheulian) 주먹도끼로 불리며, 유럽과 아프리카에서는 전체적인 석기군의 특성을 아우르는 아슬리안이라는 용어를 사용합니다. 이러한 주먹도끼는 구석기시대의 중요한 유물로, 고전적인 아슬리안 주먹도끼는 아몬드나 타원의 평면 형태를 가지고 있으며, 끝이 뾰족하고 단면은 양쪽이 볼록한 렌즈 모양을 가지고 있습니다.

주먹도끼는 골선사람(Homo erectus)이 만든 도구로, 주먹에 쥐고 사용하는 석기를 말합니다. 이 도구는 주로 돌로 만들어졌으며, 끝이 뾰족하거나 전체적으로 둥근 타원형을 가지고 있습니다. 주먹도끼는 골선사람의 불을 다루고 도구를 만들며 생활 영역을 확대하는 데 사용되었습니다. 서양에서는 아슬리안(Acheulian) 주먹도끼로 불리며, 유럽과 아프리카에서는 전체적인 석기군의 특성을 아우르는 아슬리안이라는 용어를 사용합니다. 이러한 주먹도끼는 구석기시대의 중요한 유물로, 고전적인 아슬리안 주먹도끼는 아몬드나 타원의 평면 형태를 가지고 있으며, 끝이 뾰족하고 단면은 양쪽이 볼록한 렌즈 모양을 가지고 있습니다.

Query

주먹도끼에 대해서 알려줄래? [0.41070669889450073, 0.4084071218967438, 0.40674442052841187, 0.38531559705734253, 0.3840537965297699]
threshold를 넘는 index : [0, 1, 2, 3, 4]

Passage Retrieval

=====아래 문서를 참고합니다=====

#문서0

연천 전곡리 주먹도끼

그렉 보웬(Greg L. Bowen)은 미국 캘리포니아 빅터 벌리 대학에서 고고학을 공부하다가 1974년 미군에 입대하면서 한국의 동두천에서 주한 미공군 하사관으로 근무하게 되었습니다. 그는 1978년 한탄강 유원지에 한국인 애인과 함께 산책을 갔다가 토기편을 발견하고 그 주위를 조사하면서 주먹도끼를 발견하게 되었습니다. 그때만 해도 이 주먹도끼가 동아시아 구석기 고고학에 있어 기념비적인 발견이 되리라고는 그 누구도 상상하지 못했습니다

#문서1

주먹도끼

주먹도끼의 제작자는 인류의 직계 조상인 골선사람(Homo erectus)입니다. 골선사람 이전까지의 인류 조상들은 아프리카에서 살았습니다. 그러나 지금으로부터 약 160만 년 전 쯤에 출현한 골선사람들은 불을 능숙하게 다듬으로써 비로소 아프리카를 벗어나 유럽과 아시아로까지 삶의 영역을 확대시킬 수 있었습니다. 골선사람의 화석 출토 범위와 주먹도끼가 발견되는 지역의 범위는 대체로 일치합니다

#문서2

주먹도끼

주먹도끼 -> 국적/시대:한국(韓國)-구석기(舊石器), 크기길이 23.6cm, 문화재구분정보없음, 소장품번호신수(新收) 019143 00000

#문서3

연천 전곡리 주먹도끼

주먹도끼는 형태적으로 끝이 뾰족하거나 전체적으로 둥근 타원형의 석기를 말하며, 기능적으로는 주먹에 쥐고 사용했던 석기를 말합니다

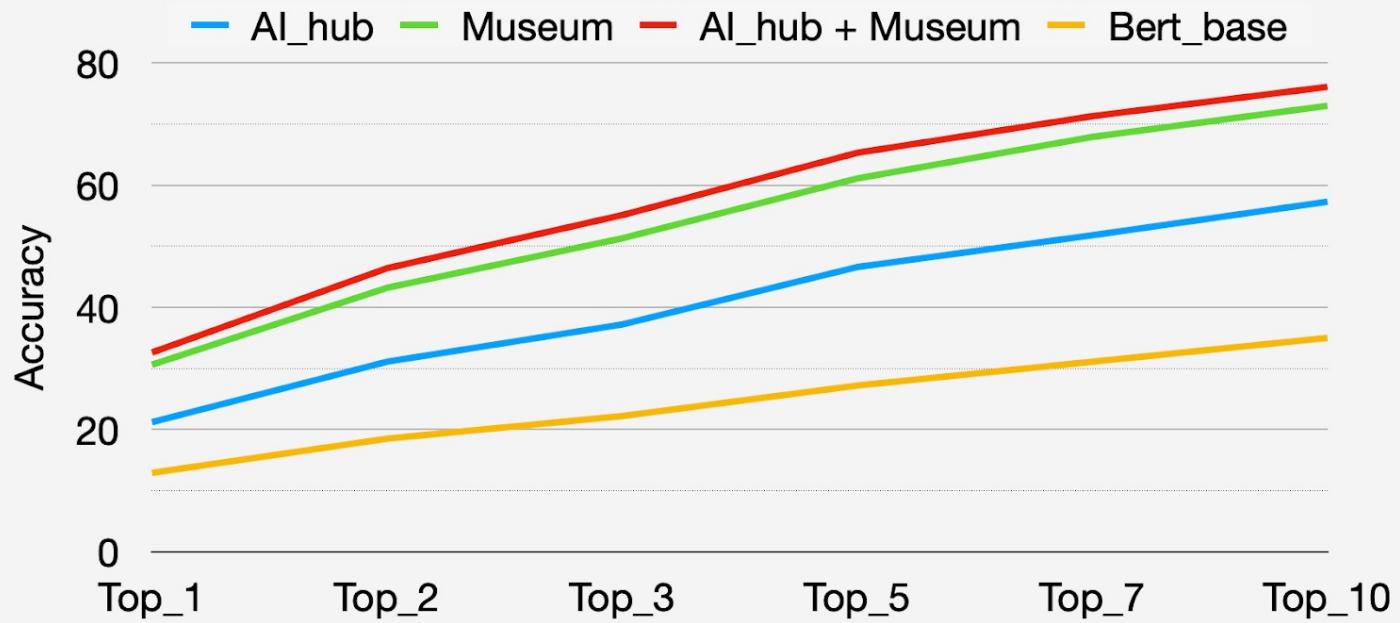
#문서4

연천 전곡리 주먹도끼

서양의 주먹도끼는 아슬리안(Acheulian)으로 불리는데, 프랑스 북서부 솜드강 강변에 위치한 생 아슬(St. Acheul)에서 다양한 석기가 확인되면서 불여진 명칭입니다. 유럽과 아프리카에서는 개별 도구를 나타내는 주먹도끼라는 용어보다는 전체적인 석기군의 특성을 아우르는 아슬리안이라는 용어를 많이 사용합니다. 서양에서 부르는 고전적인 아슬리안 주먹도끼는 아몬드나 타원의 평면 형태를 가지고 있으며 끝이 뾰족하며, 단면은 양쪽이 볼록한 렌즈 모양입니다. 석기 주변부를 따라서 날카로운 날이 있고, 양면 전체를 빼내어 만들입니다

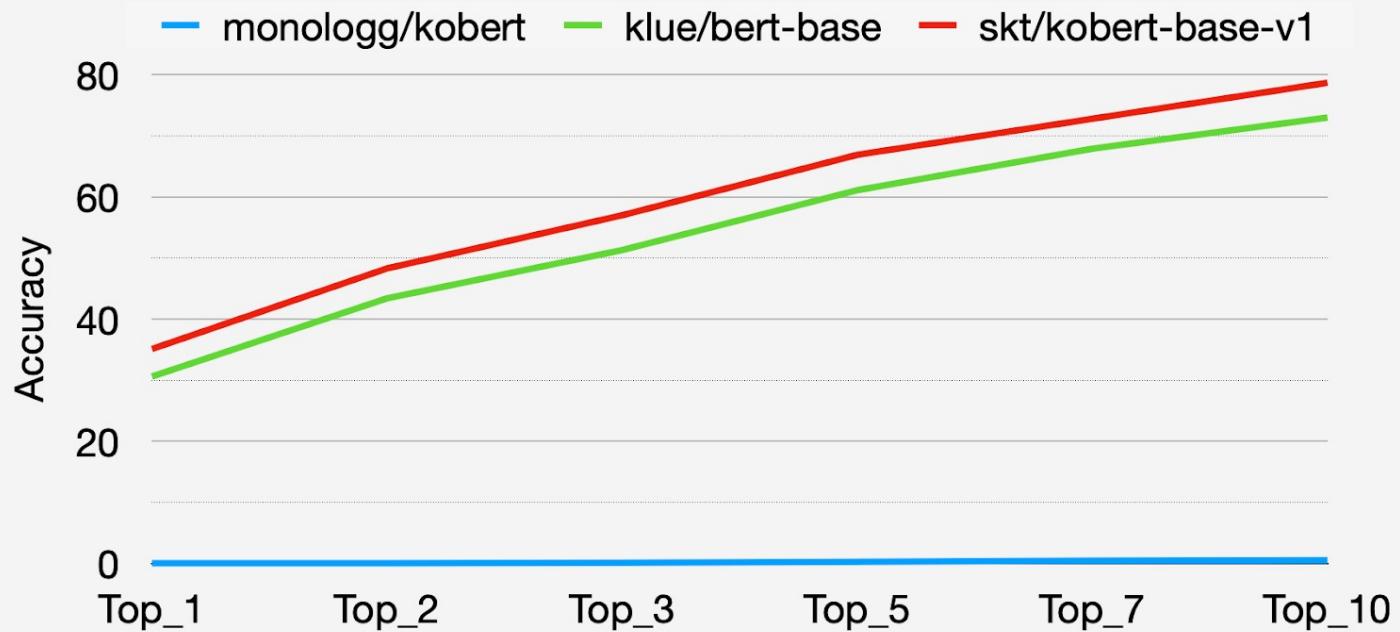
데이터별 성능평가

	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
AI_hub	21.2	31.1	37.2	46.6	51.8	57.3
Museum	30.6	43.2	51.3	61.1	67.9	73
AI_hub + Museum	32.6	46.4	55.1	65.3	71.3	76.1
Bert_base	12.9	18.5	22.2	27.2	31.1	35



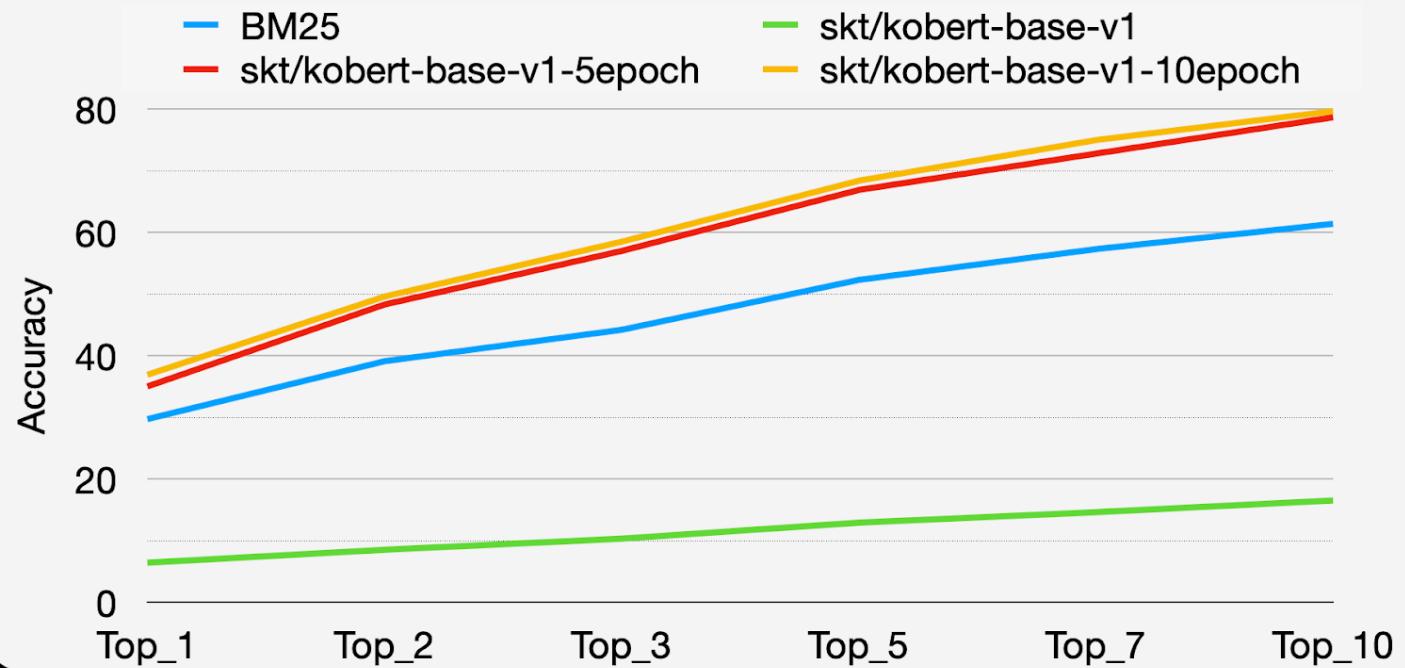
모델별 성능평가

	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
monologg/kobert	0.03	0.03	0.11	0.26	0.44	0.52
klue/bert-base	30.6	43.4	51.3	61.1	67.9	73
skt/kobert-base-v1	35.1	48.3	57	66.9	72.8	78.7



BM25 vs 싱글 인코더 vs DPR 학습

	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
BM25	29.7	39.1	44.2	52.3	57.3	61.4
skt/kobert-base-v1	6.4	8.5	10.3	12.9	14.6	16.5
skt/kobert-base-v1-5epoch	35	48.3	57	66.9	72.8	78.7
skt/kobert-base-v1-10epoch	36.9	49.6	58.5	68.4	75	79.7



최종 성능 비교 결과

Dataset (klue/bert-base)	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
AI hub	21.2	31.1	37.2	46.6	51.8	57.3
Museum	30.6	43.2	51.3	61.1	67.9	73.0
AI hub + Museum	32.6	46.4	55.1	65.3	71.3	76.1
Model (Museum Data)	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
monologg/kobert-5epoch	0.03	0.03	0.11	0.26	0.44	0.52
klue/bert-base-5epoch	30.6	43.4	51.3	61.1	67.9	73.0
skt/kobert-base-v1-5epoch	35.1	48.3	57.0	66.9	72.8	78.7
Model (Museum Data)	Top_1	Top_2	Top_3	Top_5	Top_7	Top_10
Bm25	29.7	39.1	44.2	52.3	57.3	61.4
skt/kobert-base-v1	6.4	8.5	10.3	12.9	14.6	16.5
skt/kobert-base-v1-5epoch	35.0	48.3	57.0	66.9	72.8	78.7
skt/kobert-base-v1-10epoch	36.9	49.6	58.5	68.4	75.0	79.7

Appendix. GPTs의 한계점

슈팅스타

안녕하세요! 저는 박물관 AI도슨트 슈팅스타입니다.

민무늬 토기에 대해 알려줘

민무늬 토기는 청동기시대를 대표하는 유물로, 기하학적 무늬가 가득한 토기로서, 한반도의 청동기시대를 상징하는 아이콘이라고 합니다. 민무늬 토기는 청동기시대의 대표적인 유물로 인식되며, 크기는 높이 19.4 cm이고, 소장품번호는 신수 10887입니다. 이 문서를 토대로 말하자면, 민무늬 토기는 청동기시대의 문화를 대표하는 지표로 볼 수 있습니다.

금으로 만들어진 유물 중 하나를 알려주세요.

금으로 만들어진 유물 중 하나로는 경주 노서리 215번지에서 출토된 금목걸이가 있습니다. 이 목걸이는 우연히 발견되었으며, 금제구슬 77점과 비취빛 곱은옥 1점을 연결하여 하나의 목걸이를 만들었습니다. 이 목걸이는 신라의 가장 화려한 목걸이로서, 달개와 나뭇잎모양의 달개가 화려하게 장식되어 있습니다. 또한, 묘에서는 굽은고리 귀걸이, 금은팔찌, 반지, 토기 몇 점이 발견되었으며, 묘 내부에서는 대형 항아리와 제사와 관련된 물품들이 발견되었습니다.

모나리자는 누가 그렸어?

모나리자는 레오나르도 다빈치가 그렸어요.

GPTs

You
민무늬 토기에 대해 알려줘

슈팅스타

청동기시대에 해당하는 민무늬 토기에 대해 알려드리겠습니다. 민무늬 토기는 이름에서 알 수 있듯이 무늬가 있는 토기를 의미하는데, 이는 신석기시대 토기에 비해 상대적으로 무늬가 적다는 특징을 표현하기 위한 명칭입니다. 청동기시대 초기 토기를 자세히 관찰하면, 구멍무늬, 눈금무늬, 빗금무늬 등 다양한 무늬가 표현되어 있음을 볼 수 있습니다. 이러한 무늬들은 시간성과 지역성, 그리고 집단성을 반영합니다. 예를 들어, 점토 따로 꾸며진 돌은 떠무늬 토기는 한강 유역이나 남강 유역에서, 금강 유역에서는 짧은 빛금무늬가 새겨진 토기가 주로 나타납니다.

You
금으로 만들어진 유물 중 하나를 알려주세요.

슈팅스타

금으로 만들어진 특별한 유물을 소개해 드리겠습니다. 이 유물은 삼국시대 백제 시대(4~5세기)의 청동 초두입니다. 초두는 고대에 만들어진 세 개의 다리와 긴 손잡이가 달린 금속 냄비로, 술이나 차, 약을 끓이는 데 사용되었습니다. 특히 한 점은 이 초두가 주로 청동이나 철로 만들어졌음에도 불구하고, 드물게 자기로 만들어진 예도 있습니다. 이 유물은 중국 한나라 때 처음 만들어져 삼국시대부터 우리나라에서 사용되기 시작했습니다.

You
모나리자는 누가 그렸어?

슈팅스타

모나리자는 레오나르도 다빈치가 그린 그림입니다. 이 작품은 16세기 초 이탈리아 르네상스 시대에 제작되었으며, 다빈치의 대표적인 작품 중 하나로 꼽힙니다. 현재 모나리자는 프랑스 파리의 루브르 박물관에 전시되어 있습니다.

한계점 및 아쉬웠던 점

자원

- 자원 부족과 모델의 크기로 인해 경량의 base 모델만 사용

데이터셋 비교 부족

- 질문 데이터셋의 퀄리티에 따라 성능이 좌우되는데, 현재 데이터셋은 특정 키워드에 대한 단순한 질문이 대다수
- 시간이 부족하여 여러 데이터를 실험/비교해보지 못했음

DPR 자체의 한계

- top-k개의 passage를 뽑아 답변을 하는 방식이므로
전체 passage 혹은 여러 document를 참조해야 하는 질문에는 답하기 어렵다.
예) “청동기 시대의 유물은 무엇이 있을까?”