

AI 챗봇을 활용한 온라인 강의 웹 서비스



Online Lecture Web Service Using AI Chatbot

20171612 박상현

20197132 주준하
담당교수 : 임경태

20171768 남승완

한밭대학교 컴퓨터공학과

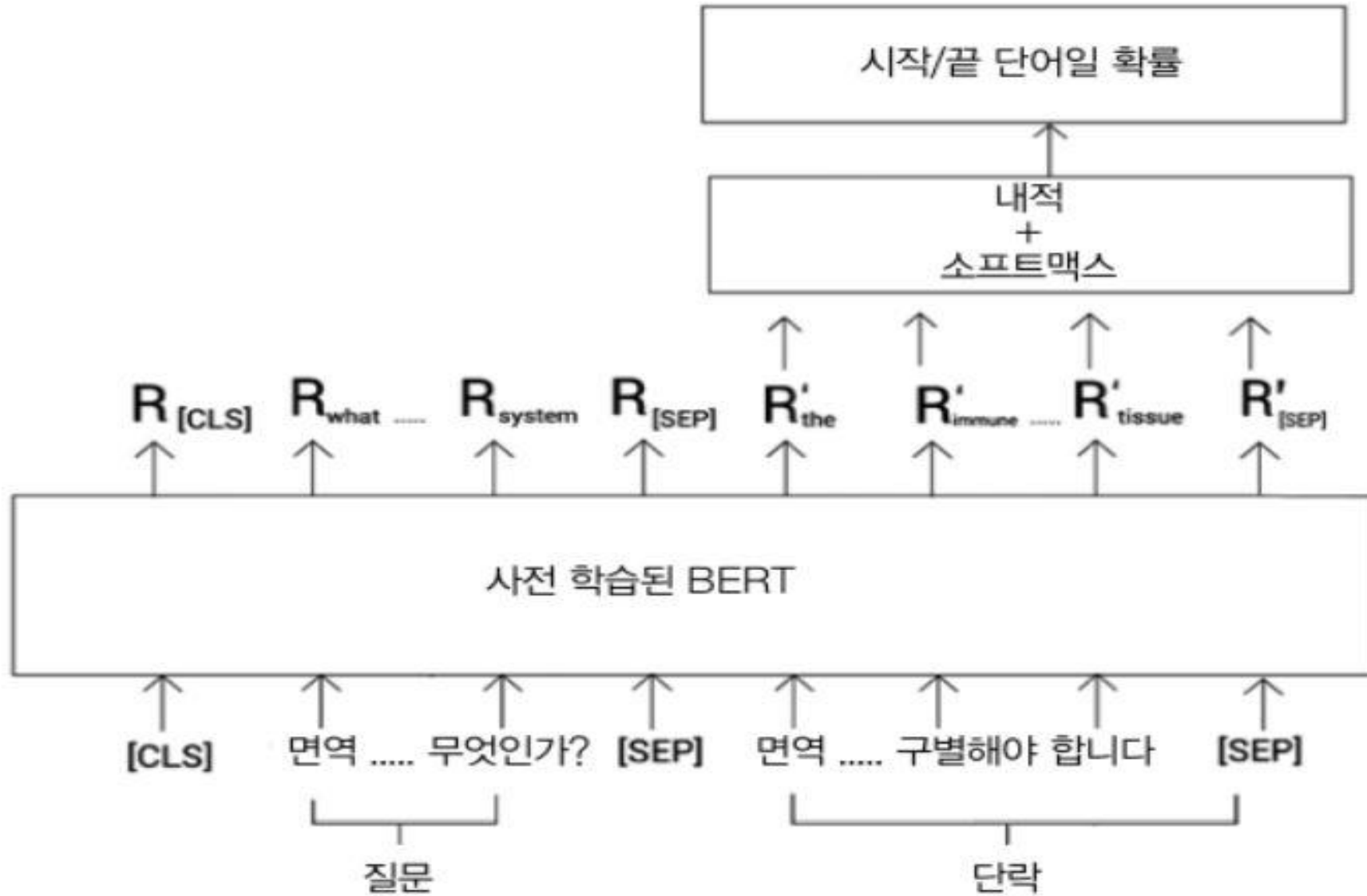


목차...

1. 프로젝트 목표
2. 초기 설계
3. 중간 설계
4. 진행상황 및 계획
5. 참고문헌

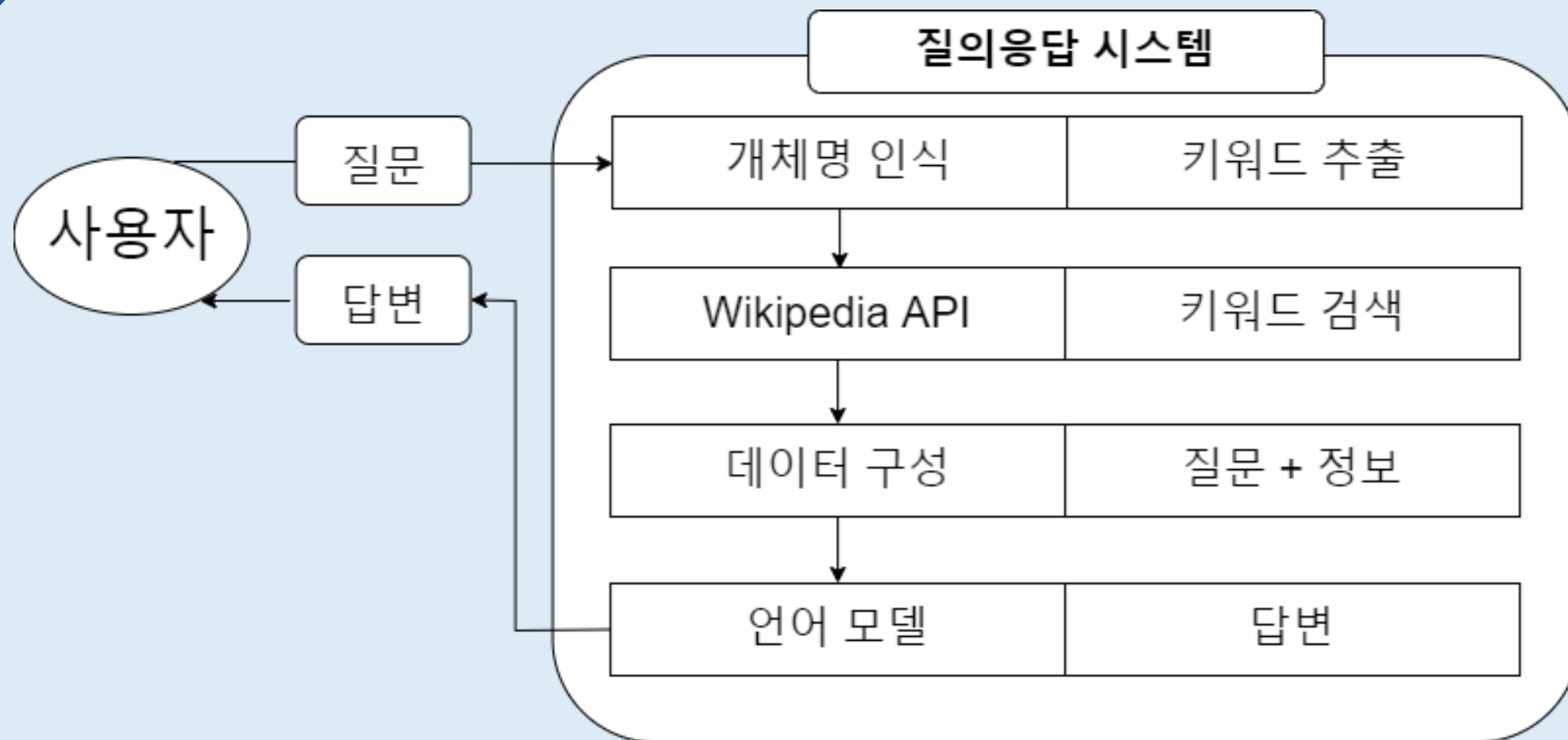
목차





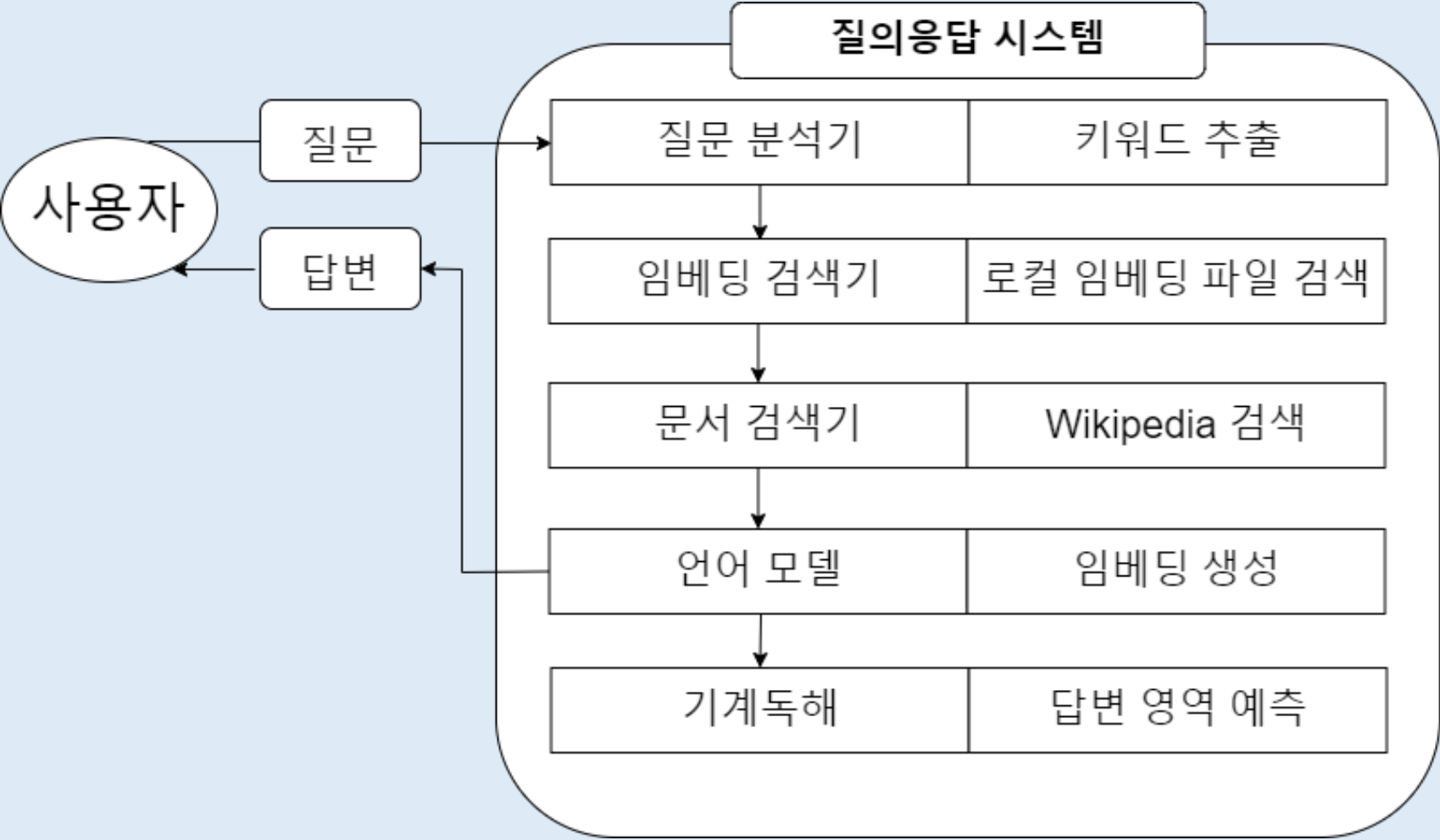


2. 초기설계



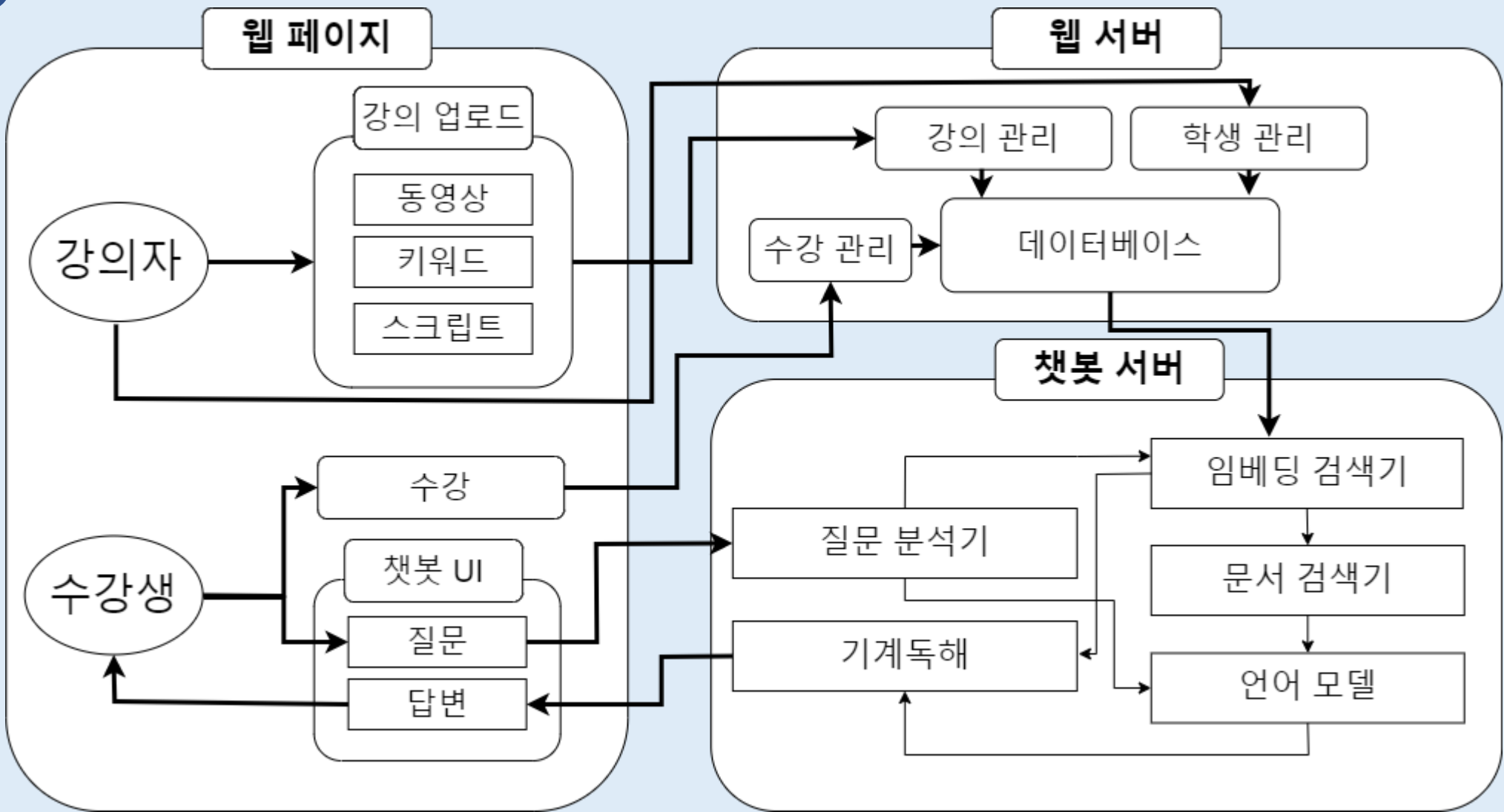


3. 중간설계





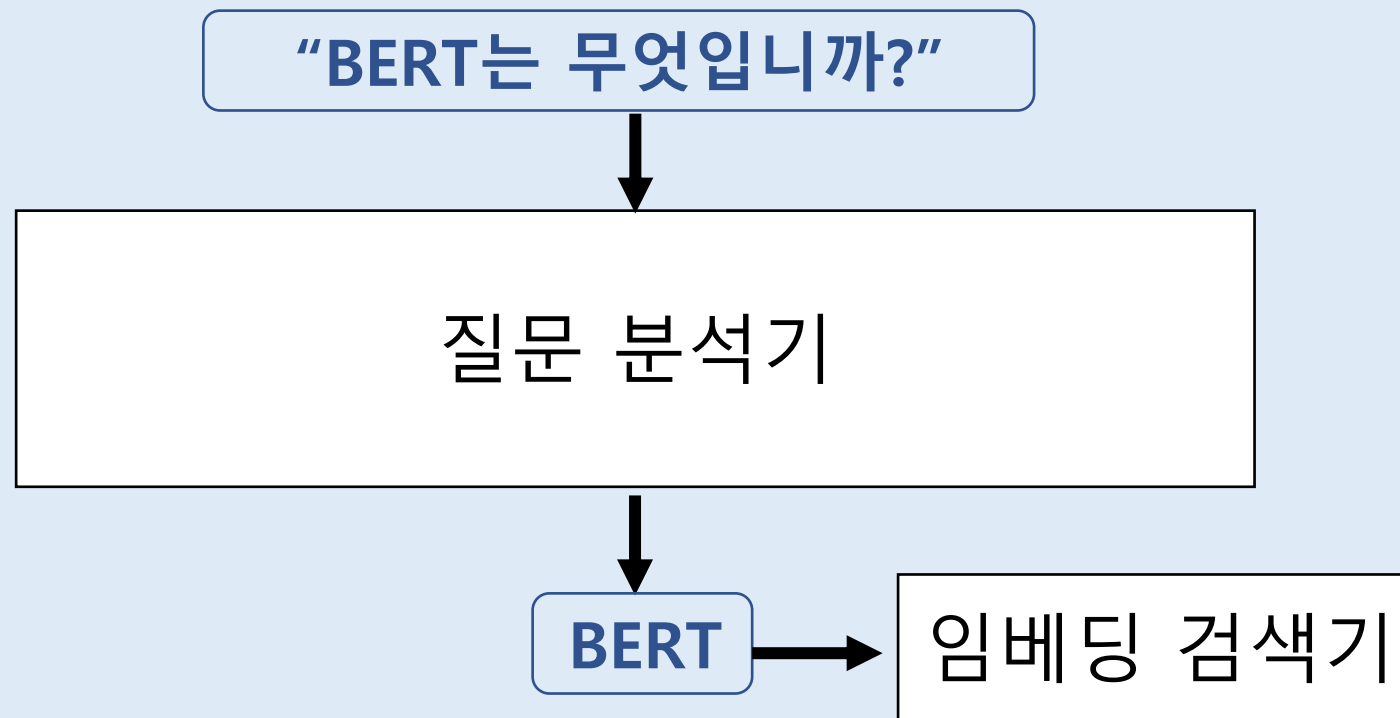
3. 중간설계





3. 중간설계-질문 분석기

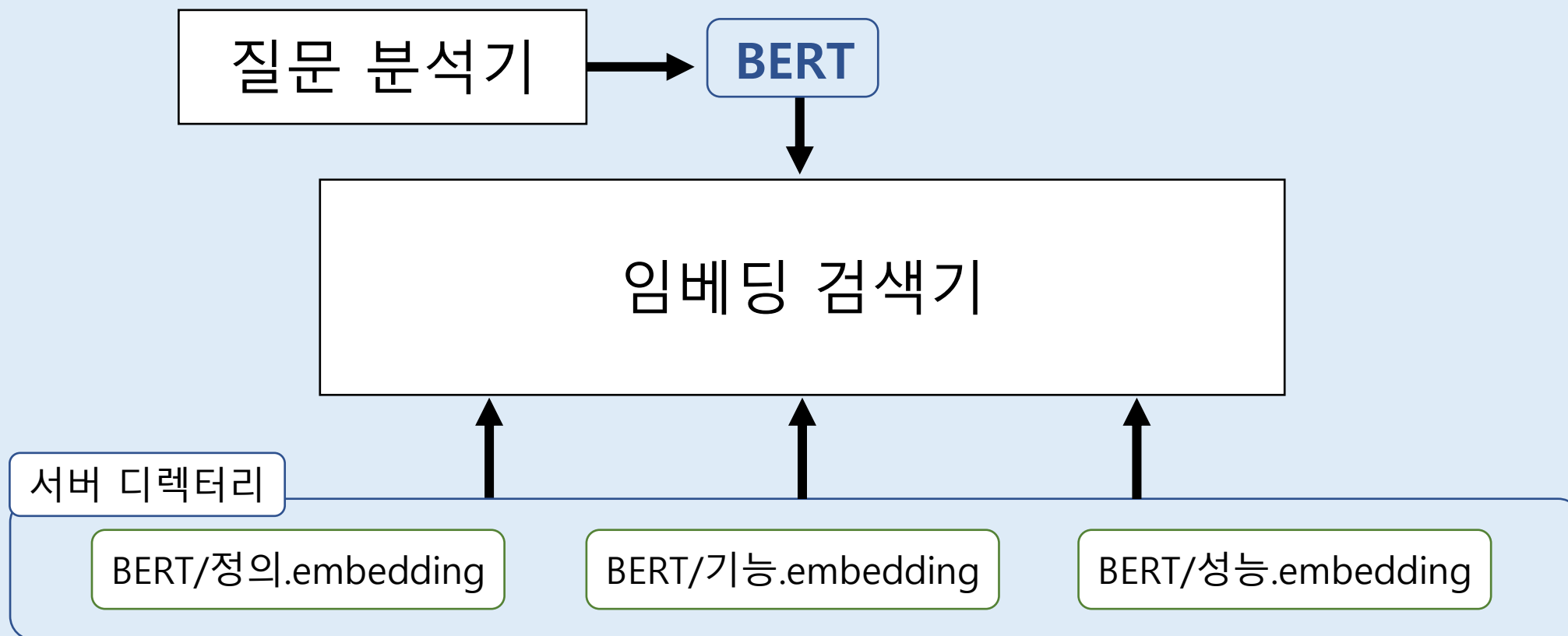
사용자 질문 분석, 주요 키워드 추출





3. 중간설계-임베딩 검색기

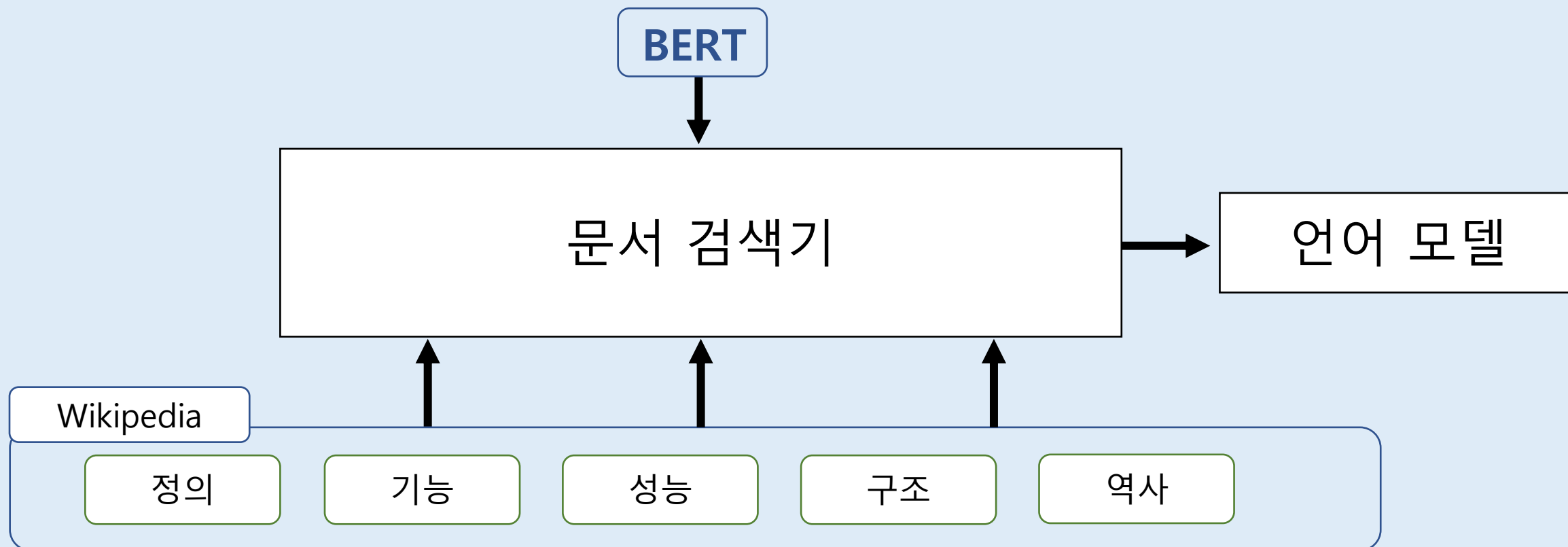
서버 내 저장된 키워드에 대한 임베딩 검색





3. 중간설계-문서 검색기

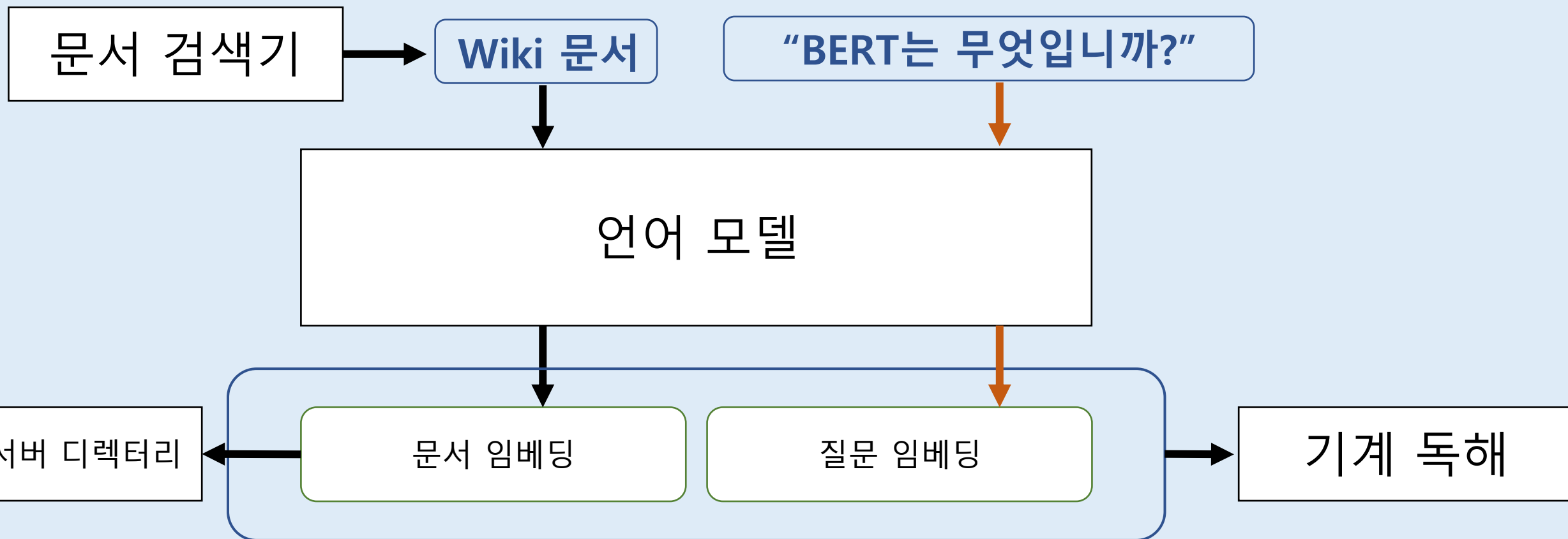
저장된 데이터가 없을 때, Wikipedia에서 키워드에 대한 정보를 검색





3. 중간설계-언어 모델

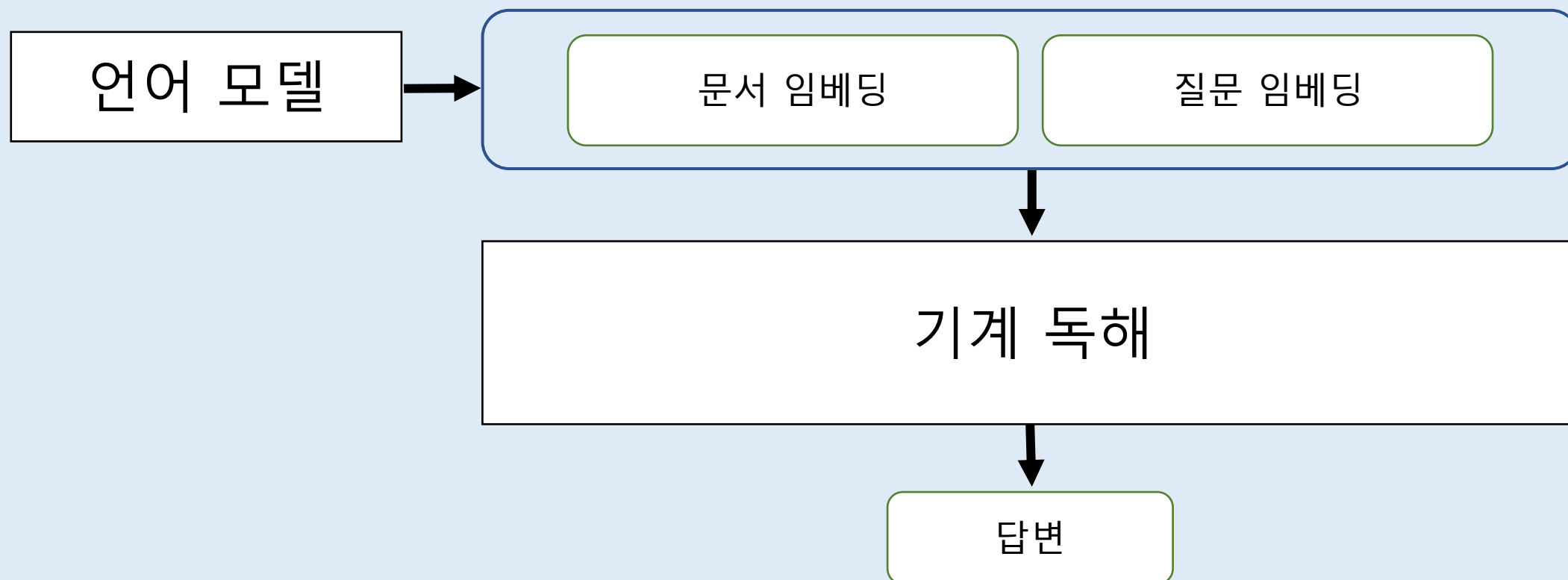
BERT 기반의 자연어 임베딩 생성 모델





3. 중간설계-기계 독해

사용자 질문에 대한 임베딩과 문단 임베딩을 이용한 답변 영역 예측





4. 진행 상황 및 계획-질문 분석기

ETRI 형태소분석 open api 활용

```
>>> print(question)
BERT는 무엇입니까?
>>> q_json['sentence'][0]['morp']
[{'id': 0.0, 'lemma': 'BERT', 'type': 'SL', 'position': 0.0, 'weight': 1.0}, {'id': 1.0, 'lemma': '는', 'type': 'JX', 'position': 4.0, 'weight': 0.111517}, {'id': 2.0, 'lemma': '무엇', 'type': 'NP', 'position': 8.0, 'weight': 0.0672403}, {'id': 3.0, 'lemma': '이', 'type': 'VCP', 'position': 14.0, 'weight': 0.0549335}, {'id': 4.0, 'lemma': '니까', 'type': 'EF', 'position': 14.0, 'weight': 0.0978259}, {'id': 5.0, 'lemma': '?', 'type': 'SF', 'position': 23.0, 'weight': 1.0}]

>>> print(question)
인공지능은 무엇입니까?
>>> q_json['sentence'][0]['morp']
[{'id': 0.0, 'lemma': '인공', 'type': 'NNG', 'position': 0.0, 'weight': 0.101377}, {'id': 1.0, 'lemma': '지능', 'type': 'NNG', 'position': 6.0, 'weight': 0.044519}, {'id': 2.0, 'lemma': '은', 'type': 'JX', 'position': 12.0, 'weight': 0.107008}, {'id': 3.0, 'lemma': '무엇', 'type': 'NP', 'position': 16.0, 'weight': 0.0687177}, {'id': 4.0, 'lemma': '이', 'type': 'VCP', 'position': 22.0, 'weight': 0.0549414}, {'id': 5.0, 'lemma': '니까', 'type': 'EF', 'position': 22.0, 'weight': 0.0978285}, {'id': 6.0, 'lemma': '?', 'type': 'SF', 'position': 31.0, 'weight': 1.0}]
```

명사(NNG), 고유명사(NNP), 외국어(SL) 추출하여 키워드 사용



4. 진행 상황 및 계획-임베딩 검색기

1. 기계 독해 방법에 따라 필요 없을 수 있다.
2. 어떤 임베딩 파일에서 답을 검색할 것인가
 - 검색하는 임베딩 파일이 많을 수록 정확도가 떨어진다.



4. 진행 상황 및 계획-문서 검색기

인공지능

챗봇

- 강인공지능(strong AI, 범용인공지능, AGI...
- 강인공지능(범용인공지능)
- 강인공지능strong AI 범용인공지능 AGI...
- 강인공지능범용인공지능
- 강인공지능에 대한 철학적인 주장과 반론
- 같이 보기
- 게임 인공지능
- 관련 서적
- 갈꿈이
- 갈꿈이 논리 프롤로그와 전문가 시스템
- 갈꿈함의 승리
- 낙관론
- 다트머스 컨퍼런스 1956년

강인공지능(strong AI, 범용인공지능, AGI)과 약인공지능(weak AI) - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

초기 인공지능 연구에 대한 대표적인 정의는 다트머스 회의에서 존 마

- 각주
- 같이 보기
- 개발
- 부작용
- 외부 링크
- 챗봇
- 한계

챗봇 - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

챗봇(chatbot) 혹은 채터봇(chatterbot)은 음성이나 문자를 통한 유
인간이 사용하는 언어를 이해해서 대화를 진행하는 자연언어처리
그리고 각본을 미리 만들고 각본에 따라서 사용자의 입력에 대한
네이버클라우드플랫폼, 카카오톡, LINE, 슬랙, 페이스북 메신저, 스



4. 진행 상황 및 계획-BERT

vocab.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

можат
D"p
byens
podobnie
##ültür
plura
##기는
obejmuje
señal
операция
##ᄅᆞᆫ
##지 고
は3
##udou
incluem
...

Ln 68829, Col 4100%Unix (LF)UTF-8

vocab.korean_morp.list - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

있/VX_ 2984347
있/VA_ 2766956
년/NNB_ 2692800
하/VV_ 2640633
ㄴ 다/EF_ 2593343
하/XSA_ 2417945
았/EP_ 2248550
일/NNB_ 2208697
은/ETM_ 2165340
과/JC_ 2080158
계/EC_ 2045603
지/EC_ 2021309
기/ETN_ 1982615
1/SN_ 1747264
등/NNB_ 1741349
ㅈ/VCA_ 1691990

Ln 8, Col 4100%Unix (LF)UTF-8



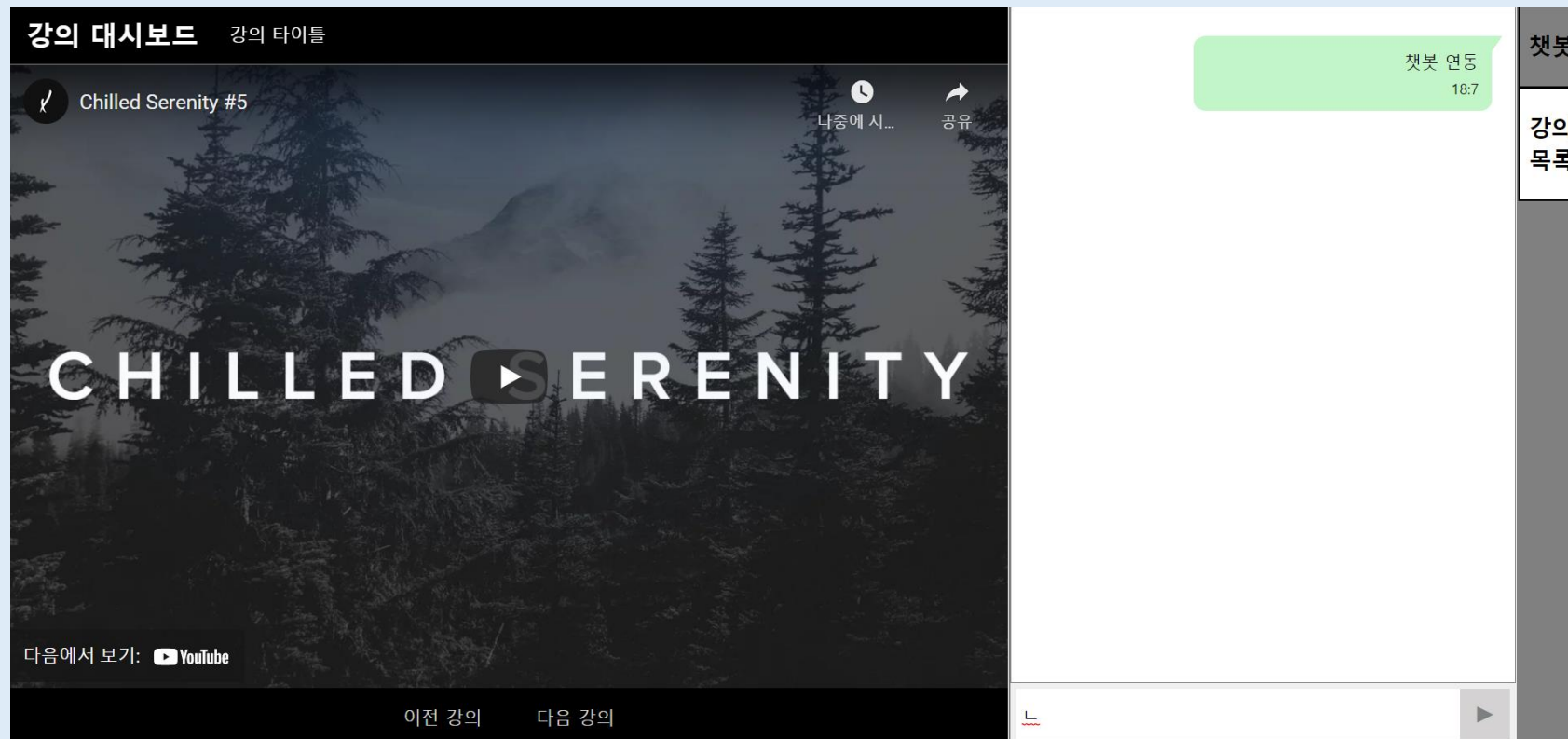
4. 진행 상황 및 계획-기계독해

기계독해 방법 2가지

1. 질문 벡터와 문서 내 문장들의 벡터의 유사도를 측정하여 가장 높은 유사도를 가진 문장을 답변으로 사용 – 간단하지만 정확도가 낮은 방법
2. 생성된 각 토큰의 임베딩 벡터와 시작 벡터, 종료 벡터(학습되는 벡터) 내적을 구하고 소프트맥스를 취해서 답변의 시작 토큰과 종료 토큰을 구해서 답변 생성 – 기본 fine tuning 모델에서 사용하는 방법.



4. 진행 상황 및 계획-웹



강의 페이지 구현과
챗봇 UI구현

이후 챗봇 서버와 연동.



5. 참고 문헌

1. KorQuAD를 활용한 한국어 오픈도메인 질의응답 시스템 - 부산대학교 전기전자컴퓨터공학과 - 조상현, 김민호, 권혁철
2. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding - Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language
3. korBERT – ETRI
4. BERT QA fine tuning – 구글 BERT의 정석



감사합니다