

데이터 수집 및 저장 데이터 베이스 설계서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터베이스 설계서
- 제출 일자 : 2025. 02. 10
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN05-final-3Team>
- 작성 팀원 : 허상호

개요	<ul style="list-style-type: none">• 소개• 시스템 개요
구조	<ul style="list-style-type: none">• 시스템 아키텍처• 요구사항 매트릭스
설계	<ul style="list-style-type: none">• 데이터 설계• 설계 근거• ERD

소개	<ul style="list-style-type: none">- 목적<ul style="list-style-type: none">- 이 데이터베이스는 사용자가 크롬 확장 프로그램을 통해 구글 독스 문서를 기반으로 질문과 답변(QnA)을 수행할 수 있도록 데이터를 저장하고 관리하기 위해 설계됨.- 범위<ul style="list-style-type: none">- 사용자 로그인 정보, 문서 업로드 정보, 세션 관리, QnA 기록 등을 포함하며, 확장 프로그램의 원활한 작동 및 무결성을 지원함.								
시스템 개요	<ul style="list-style-type: none">- 시스템 역할<ul style="list-style-type: none">- 정상적인 경로(로그인 > 구글 docs > side panel open)로 들어온 사용자만 질문을 할 수 있도록 함.- 사용자가 확장 프로그램을 통해 질문하면, LLM 모델의 답변(QnA)과 함께 질문 내용을 저장함.- 데이터베이스는 사용자 정보, 업로드한 pdf 문서 정보, QnA 기록, 세션 데이터를 저장하고 관리함.- 주요 기능<ul style="list-style-type: none">- 사용자 인증 및 로그인 시간 기록- 업로드된 문서 관리- 세션별 QnA 기록 저장- QnA 이력 추적 및 데이터 소스 관리								
시스템 아키텍처	<ul style="list-style-type: none">- 데이터베이스 구조<ul style="list-style-type: none">- 관계형 데이터베이스(RDBMS)로 설계, 주요 테이블은 다음과 같음:<ul style="list-style-type: none">- member_tbl: 사용자 정보를 저장- pdf_tbl: 업로드된 문서 정보를 저장- session_tbl: 사용자의 세션 정보를 관리- qna_tbl: QnA 기록을 관리- 적용한 데이터베이스<ul style="list-style-type: none">- AWS의 RDS를 기반으로 MySQL DB를 적용함.- 문서의 경우 chromaDB에 벡터화 하여 저장함.- MySQL에서는 문서의 이름만 저장함.								
요구사항 매트릭스	<div>요구사항 및 관련 테이블</div> <table><tr><th>요구사항</th><th>관련 테이블</th></tr><tr><td>사용자 로그인 정보 저장</td><td>member_tbl</td></tr><tr><td>업로드된 PDF 및 구글 독스 정보 관리</td><td>pdf_tbl</td></tr><tr><td>세션별 상호작용 추적</td><td>session_tbl</td></tr></table>	요구사항	관련 테이블	사용자 로그인 정보 저장	member_tbl	업로드된 PDF 및 구글 독스 정보 관리	pdf_tbl	세션별 상호작용 추적	session_tbl
요구사항	관련 테이블								
사용자 로그인 정보 저장	member_tbl								
업로드된 PDF 및 구글 독스 정보 관리	pdf_tbl								
세션별 상호작용 추적	session_tbl								

데이터 설계

QnA 기록 저장 및 검색		qna_tbl	
<div>- 테이블 설명</div> <div>- member_tbl</div> <div>- 사용자 이메일과 로그인 시간을 저장함.</div> <div>- 기본 키: user_email</div>			
컬럼명	데이터 타입	설명	NN(not null)
user_email	VARCHAR(40)	사용자 이메일 (PK)	True
login_time	DATETIME	마지막 로그인 시간	True
<div>- pdf_tbl</div> <div>- 사용자가 업로드한 PDF 문서 정보를 저장함.</div> <div>- 기본 키: pdf_id</div> <div>- 외래 키: user_email + docs_id</div>			
컬럼명	데이터 타입	설명	NN(not null)
pdf_id	INT	문서 ID (PK)	True
user_email	VARCHAR(40)	사용자 이메일 (FK)	True
docs_id	VARCHAR(100)	구글 독스 문서 ID	True
file_name	VARCHAR(100)	파일 이름	True
file_time	DATETIME	파일 업로드 시간	True
<div>- session_tbl</div> <div>- 사용자의 email과 google docs id를 기반으로 UUID5 기반 고유한 session id를 생성하여 사용자 별 환경을 분리함.</div> <div>- 기본 키: user_email + docs_id (복합키)</div>			
컬럼명	데이터 타입	설명	NN(not null)
user_email	VARCHAR(40)	사용자 이메일 (FK)	True
docs_id	VARCHAR(100)	구글 독스 문서 ID	True
session_id	VARCHAR(36)	UUID5 기반 세션 ID	True

	<ul style="list-style-type: none"> - qna_tbl - QnA 기록과 관련된 정보를 저장함. - 기본 키: qna_id - 외래 키: user_email + docs_id, session_id - 답변이 반환되지 않는 오류에 대비하여 answer의 NN은 False로 설정 		
	컬럼명	데이터 타입	설명
	qna_id	INT	QnA ID (PK)
	user_email	VARCHAR(40)	사용자 이메일 (FK)
	docs_id	VARCHAR(100)	구글 독스 문서 ID (FK)
	session_id	VARCHAR(36)	세션 ID (FK)
	question	VARCHAR(1000)	질문 내용
	answer	TEXT	답변 내용
	chat_option	VARCHAR(20)	채팅 옵션
	ask_time	DATETIME	질문 시간
	source	BLOB	질문 관련 추가 자료
설계 근거	<ul style="list-style-type: none"> - 데이터 무결성 보장 - 외래 키(FK)를 활용하여 테이블 간 관계를 명확히 정의함. - 확장성 고려 <ul style="list-style-type: none"> - 다수의 사용자가 동시에 작업할 수 있도록 각 세션별로 데이터를 분리하고, 대량의 QnA 데이터를 저장할 수 있도록 설계함. - 보안 <ul style="list-style-type: none"> - 사용자 이메일을 주요 식별자로 사용하며, 민감한 데이터는 암호화 또는 추가 보안 조치를 적용할 수 있음. 		

ERD

