* 과제 내용
* 대학생들이 사용하는 강의자료, 족보의 학습 자료를 자동으로 분류하여 수업별로 자동 폴더 구조를 생성하고 자료 내 핵심 키워드를 태깅하여 관리하는 시스템을 구축한다. 또한 분류된 문서를 기반으로 하는 RAG 기반 챗봇 시스템을 만들어 응답과 함께 문서 위치를 답할 수 있게 한다.
* 과제 목표
* 학부생들이 듣는 수업의 강의자료, 과제, 족보 등을 수업별로 자동 분류 및 폴더를 생성
* 위 자료들을 토대로 RAG 기반 질의응답을 하여 정보 탐색 및 활용 시간을 단축함
* 과제 수행 방법

# 단계: 데이터 수집 및 전처리

* + 다양한 파일 포맷(pdf, docx, ppt)등에서 텍스트 데이터를 추출함
  + 문서 내용을 정제 및 전처리하여 텍스트 데이터로 변환 후 키워드 및 주제 정보 산출

# 단계: 문서 분류, 태킹 및 자동 폴더 구조 생성

# 텍스트 데이터를 벡터 임베딩을 통해 문서의 특징을 벡터화하고, 사전 정의된 기준에 따라 문서를 분류

# 키워드 추출 알고리즘을 사용해 각 문서에 대해 핵심 태그를 생성.

# 분류 결과에 따라 폴더 구조를 생성하고 파일을 해당 폴더로 이동함

# 단계: RAG 기반 질의응답 시스템 구축

# 사용자의 자연어 질문을 Embedding 벡터로 변환하고, 벡터 간 유사도 계산을 통해 관련 문서를 검색

# LLM 기반으로 질문에 대한 요약 및 답변 생성

# 단계: 테스트 및 성능 평가

* + 실사용 환경에서의 성능 테스트(검색 정확도, 응답 시간 등)
  + 피드백을 바탕으로 시스템 개선 및 튜닝
* 기대효과 및 활용분야
* 자동 분류 및 태깅 시스템을 통해 흩어진 문서들을 체계적으로 정리하여 자료 관리의 효율성 향상
* RAG 기반 질의응답 기능을 통해 자료 검색에 소요되는 시간을 크게 단축하고, 학습 효율을 높인다.