

def learning()

개발 강의 추천 서비스 챗봇



플레이데이터 AI 22기 최종 프로젝트 3팀

목차



Intro

- 팀 소개
- 프로젝트 배경
- 기획 단계 산출물 - 요구사항정의서, WBS

프로젝트 개요

- 플로우 차트
- 시스템 구조
- 시연 영상 및 실제 화면
- 사용 기술셋

결론

- 추후 업데이트 사항
- 트러블슈팅 및 배운점
- 후기

팀 소개



조소영

PM, 추천 모델 개발



고영현

웹 크롤러 개발 및 데이터 수집,
DB 관리



양민호

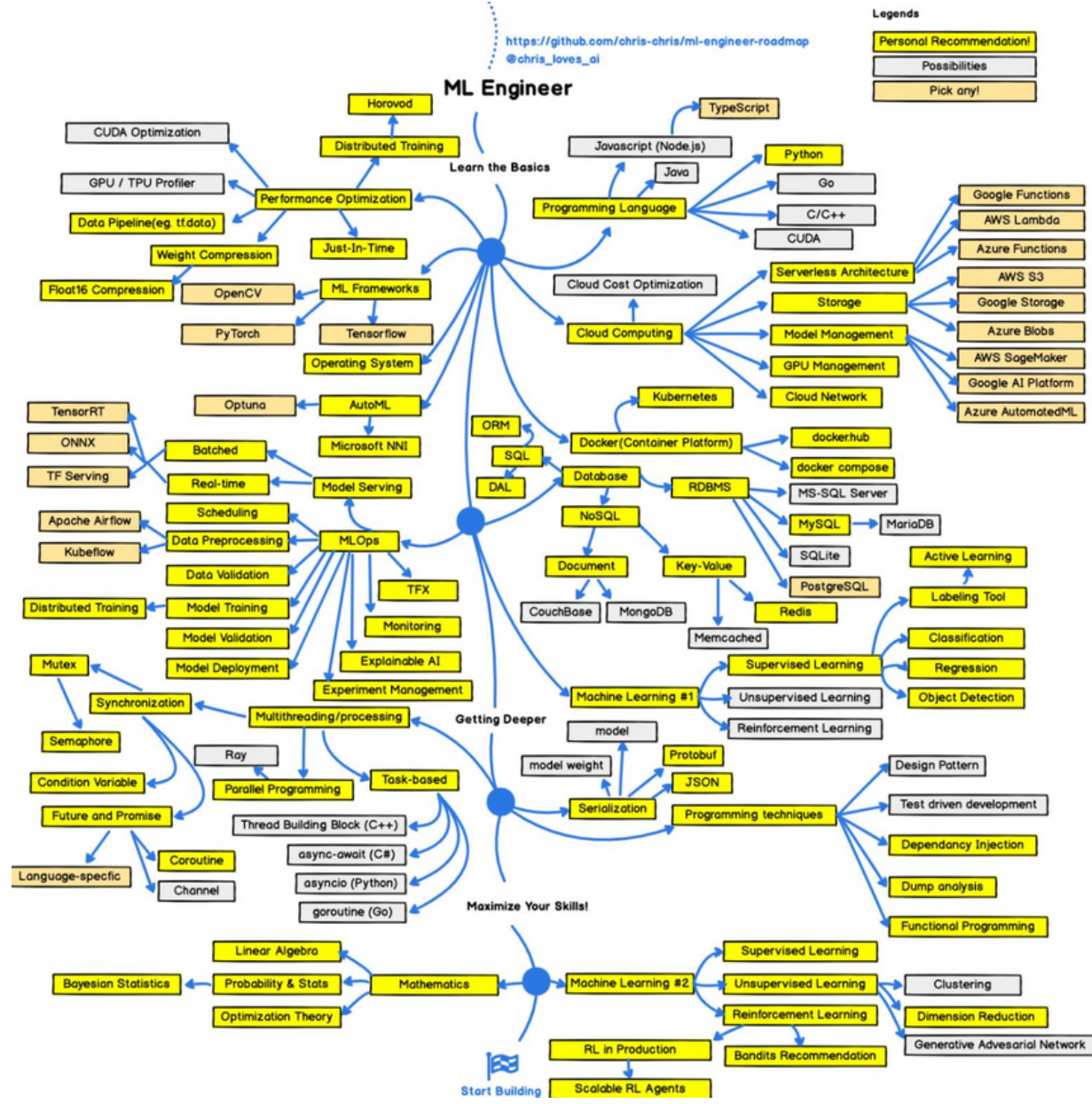
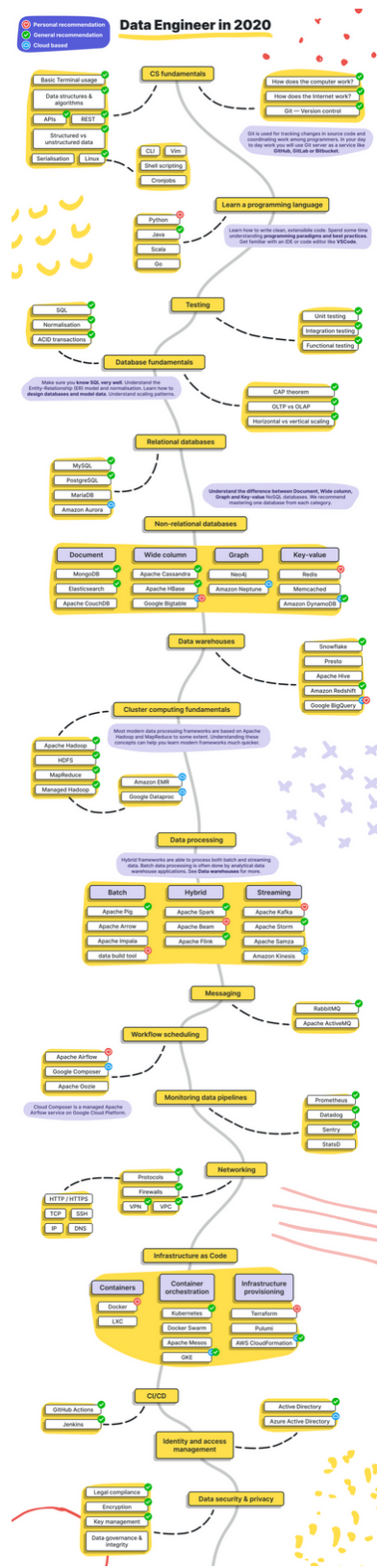
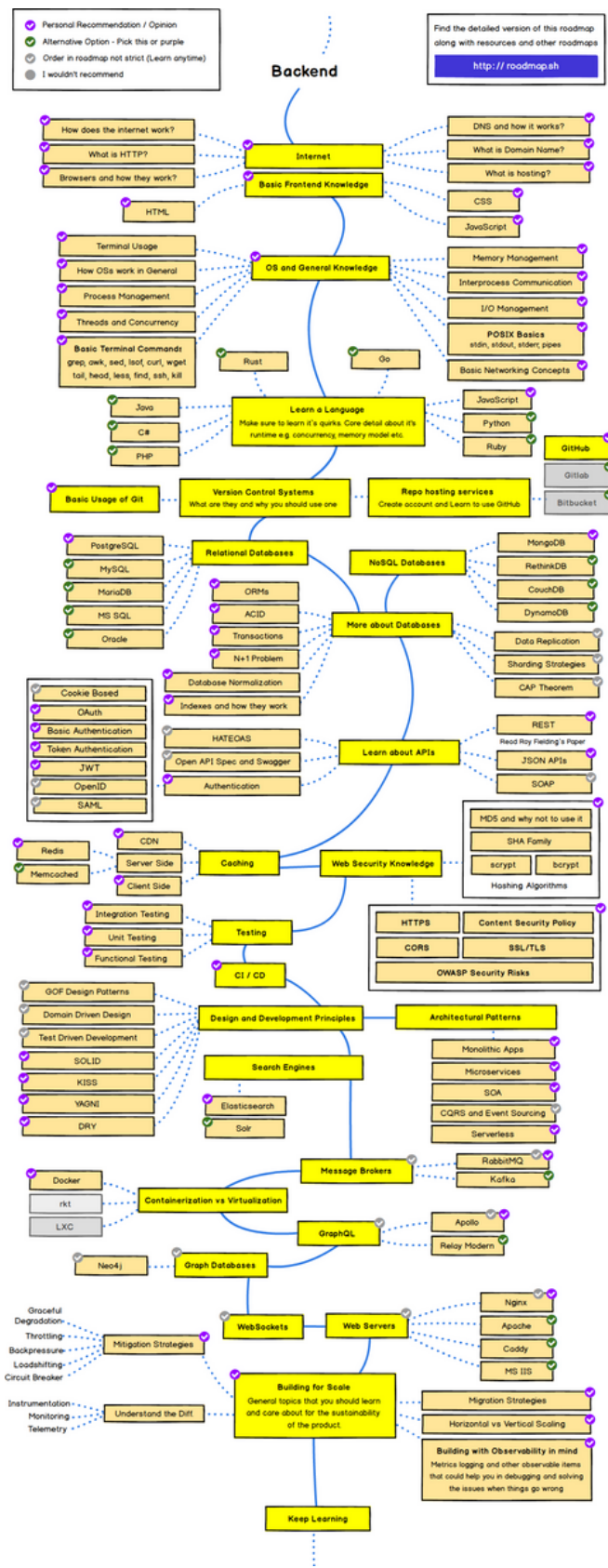
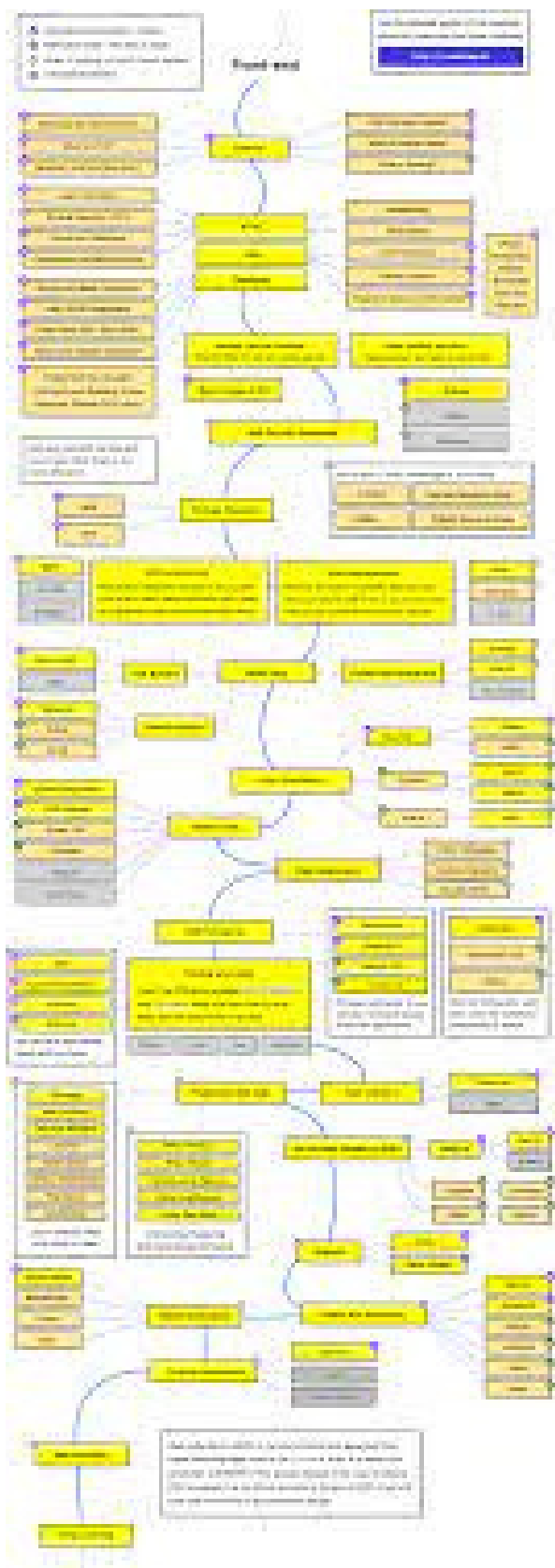
챗봇 백엔드 서버 및 기능 개발



프로젝트 배경

부트캠프 != 공부 끝

개발자에게 배움은 끝이 없는 것



분야별 로드맵

초보 개발자를 위한 필요한 강의를 간편하게 추천 받아볼 수 있는 챗봇 서비스



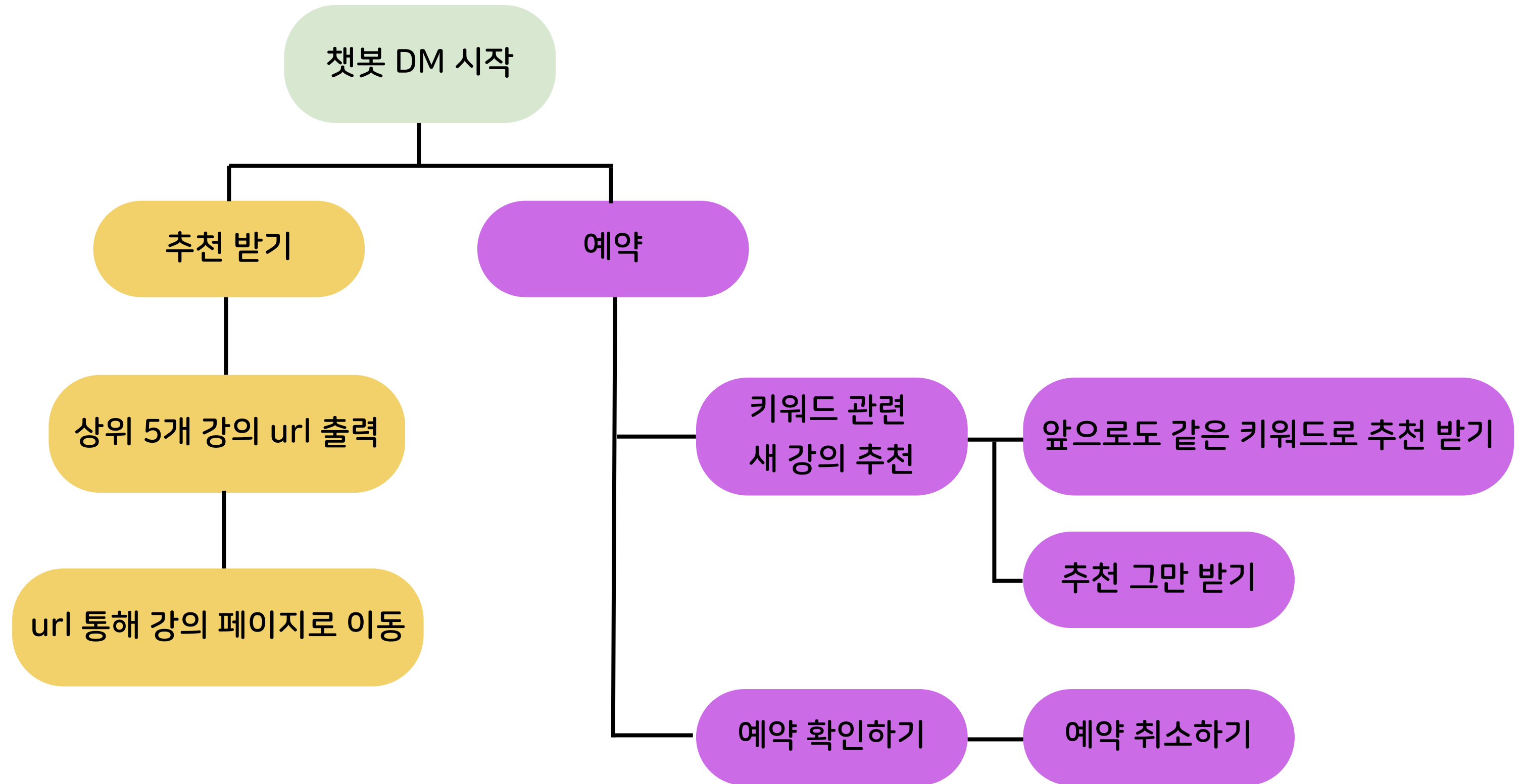
요구사항정의서

구분	요구사항	
	요구사항명	세부 요구사항
db	데이터 수집 크롤러 - 1순위 데이터	url, uuid, 강의명 키워드, 강의명, 카테고리 키워드, 카테고리, "핵심 #키워드 (분석된 키워드)", 핵심 #키워드, 리뷰 키워드, 리뷰, 강의 소개 키워드, 강의 소개, 강의 id
db	데이터 수집 크롤러 - 2순위 데이터	평점, 수강생 수, 강사명, 소요 시간, 난이도, 가격
db	데이터 수집 크롤러 - 3순위 데이터	찜 개수, 강사 소개 키워드, 강사 소개
db	데이터 전처리	특수문자 제거, 범주화, 형변환 등 필요한 1차 전처리
db	데이터 전처리 - 형태소 분석기	입력값을 형태소 분석기에서 토큰 형태로 나누고 분석에 필요한 키워드만 남김
db	데이터 적재	db에 전처리 된 데이터 적재
db	크롤러 자동화	스케줄러를 걸어서 주기적으로 크롤러가 자동으로 작동하여 강의 정보를 수집할 수 있게끔 함
추천 시스템	TF-IDF 유사도 모델	수집된 강의 데이터와 입력값의 TF-IDF, 코사인 유사도
추천 시스템	Doc2Vec 유사도 모델	수집된 강의 데이터와 입력값의 Doc2Vec, 코사인 유사도
추천 시스템	랭킹 함수	유사도에 따라 각 강의에 랭크 점수를 부여해서 랭크 상위 5개 강의를 출력하는 함수
백엔드	디스코드 챗봇 연동	discord.py 라이브러리를 사용해서 디스코드 챗봇 생성 및 연결
백엔드	사용자 입력 처리	챗봇과의 DM에 입력된 사용자 요청 값을 받아옴
백엔드	데이터 전처리 - 형태소 분석기	입력값을 형태소 분석기에서 토큰 형태로 나누고 분석에 필요한 키워드만 남김
백엔드	강의 추천	추천 모델에 토큰화된 입력값을 입력하고 결과로 추천 강의 id 5개를 받아서 사용자에게는 해당 id의 url을 출력해줌
백엔드	강의 추천 예약	원하는 적절한 강의를 현재 없을 경우 '예약' 키워드를 사용해서 강의 추천을 예약해놓고 나중에 새로 강의를 업로드 되면 강의 url을 보내줌
백엔드	예약 목록 보기	예약어 리스트를 보여줌
백엔드	예약 취소	예약어 리스트에서 취소하고 싶은 항목을 취소

WBS

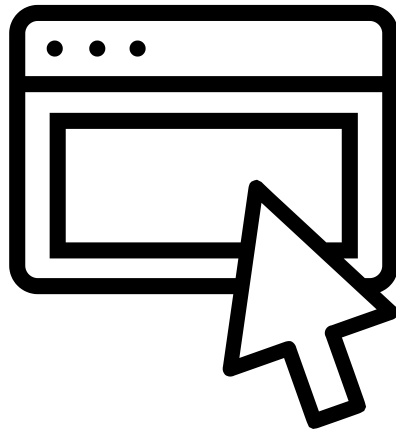
[illegible]

플로우 차트



시스템 구조 - DB

input



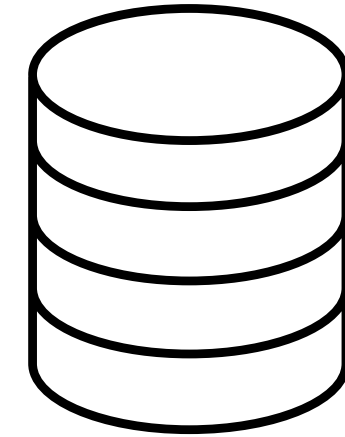
강의 사이트에서 강의명, 카테고리, 평점, 리뷰 등을 크롤링



Komoran
형태소 분석기에서
중요한 키워드만 추출

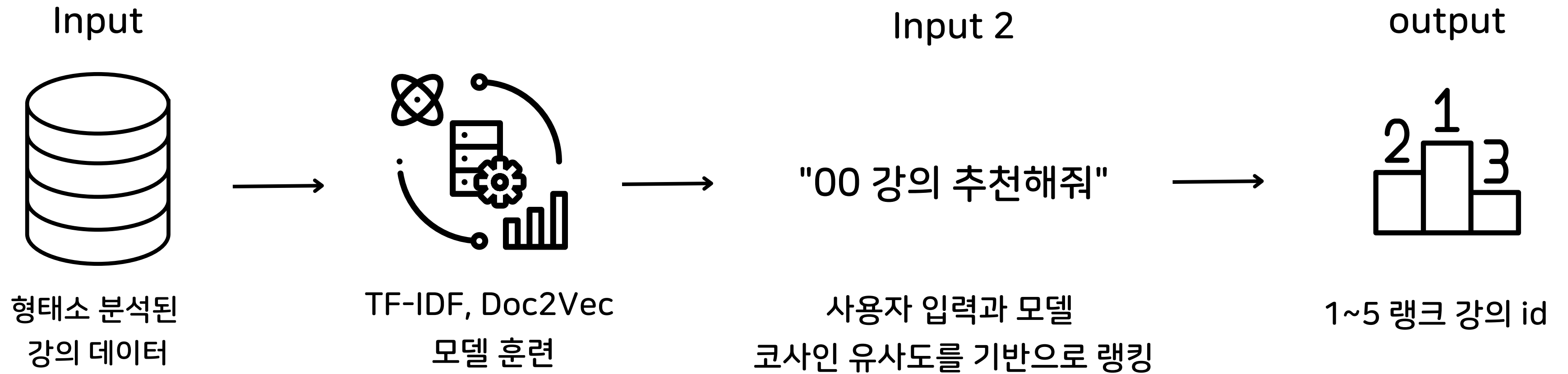


output

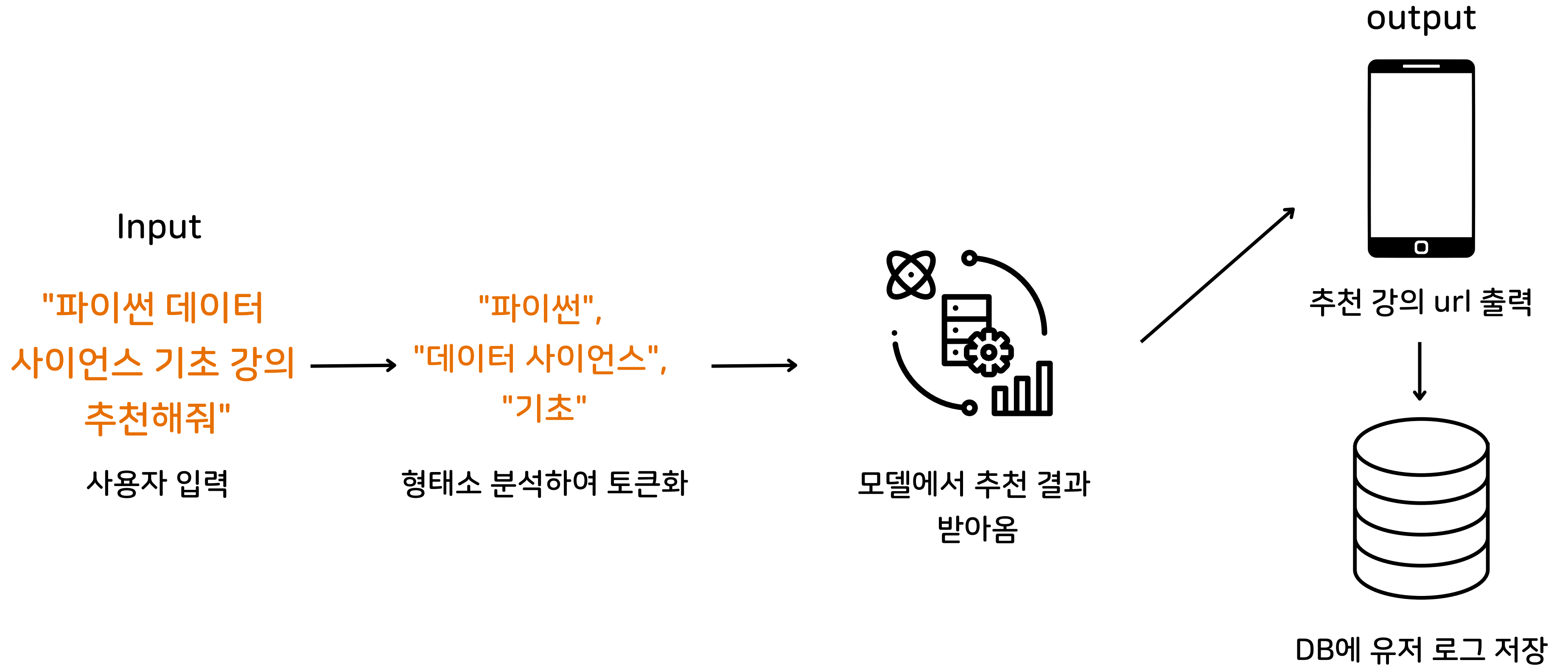


형태소 분석된 데이터
db에 저장

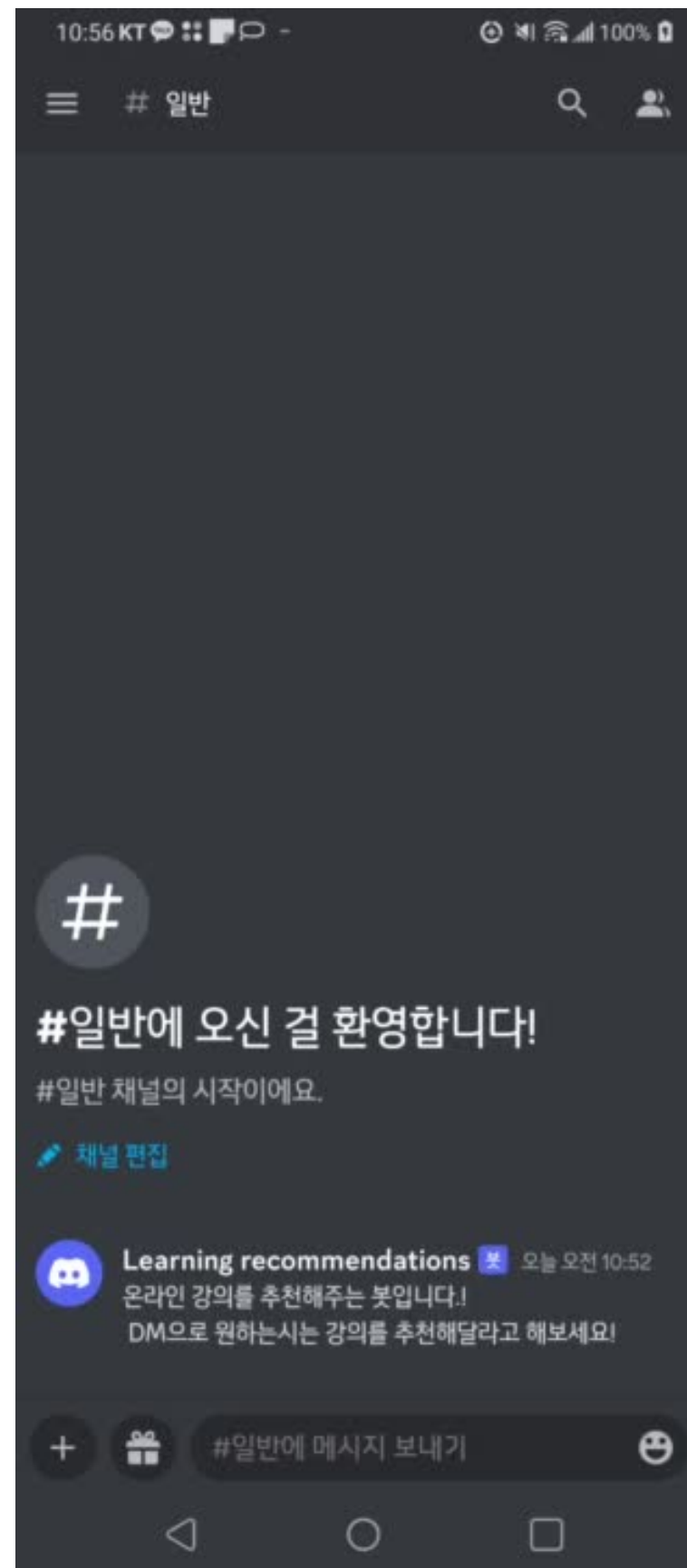
시스템 구조 - 추천 모델




시스템 구조 - 백엔드




시연 영상




실제 화면

 syja6658 2023.01.04. 오후 12:04
파이썬 강의 추천해줘

 Learning recommendations 📧 2023.01.04. 오후 12:04
<https://www.inlearn.com/course/파이썬-시작-코랩>
<https://www.inlearn.com/course/파이썬-활용7-머신러닝>
<https://www.inlearn.com/course/파이썬-활용7-머신러닝>
<https://www.inlearn.com/course/초보자-파이썬-입문>
<https://www.inlearn.com/course/입문-파이썬-기초>

인프런

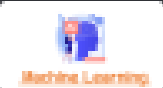
파이썬 시작하기, Colab으로 배우는 핵심 강의 1 - 인프런 | 강의



딥러닝, 사물인터넷 등 실전 프로젝트를 수행하기 위해 파이썬 프로그래밍 실전 능력을 키우고자 하는 강의입니다. 바쁜 여러분의 시간을 절약하기 위해 설치를 최소화하고 반복해서 핵심 내용을 익히도록 하나의 강의에 모든 것을 담았습니다. - 강의 소개 | 인프런...

인프런

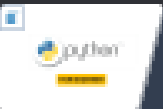
[무료] 파이썬 무료 강의 (활용편7) - 머신러닝 - 인프런 | 강의



드디어 나왔습니다. 세상에서 가장 쉬운 머신러닝 강의 무료인데 뭘 망설이세요? 어서 '학습하기' 버튼을 꾸욱 눌러보세요. - 강의 소개 | 인프런...

인프런


코딩 초보자를 위한 파이썬 입문 강의 - 인프런 | 강의



완전 초보자들과 코딩을 처음 접하는 사람을 위한 파이썬 입문 강의. - 강의 소개 | 인프런...

인프런

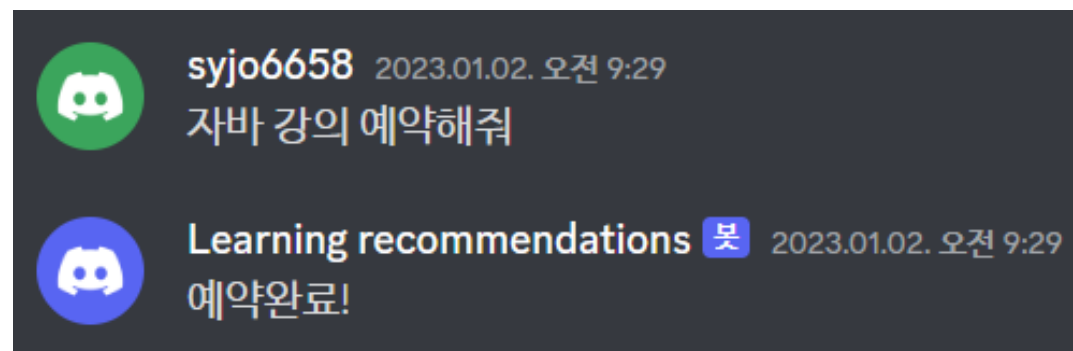
[무료] 입문자를 위한 파이썬 기초 강의 - 인프런 | 강의



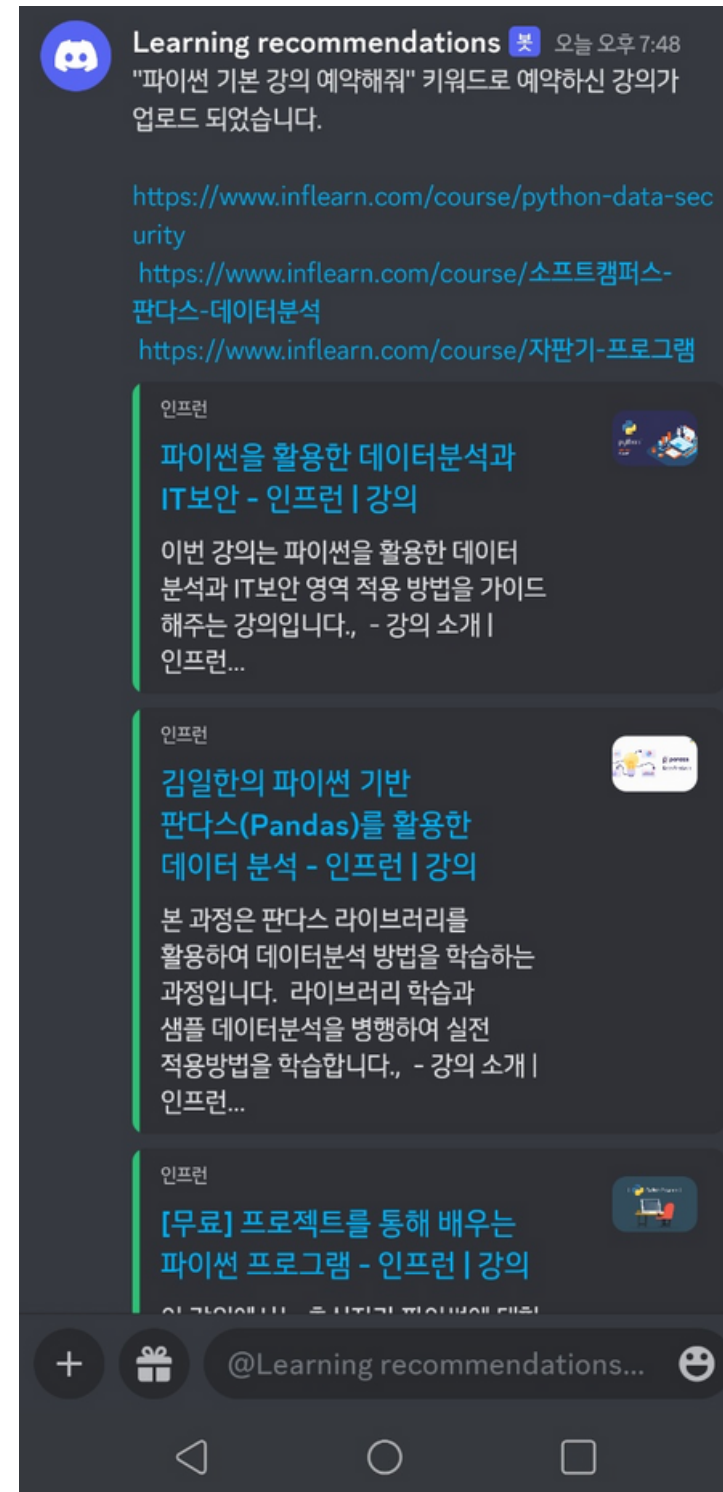
프로그래밍 언어 파이썬의 주요 개념 및 기본 문법을 소개합니다. 아무런 관련 지식을 보유하지 않은 입문자들도 소화할 수 있도록 만든 부담 없고 친절함 강의입니다. - 강의 소개 | 인프런...

강의 추천
입력 받은 키워드를 기준으로
적합한 강의 5개를 보여줌

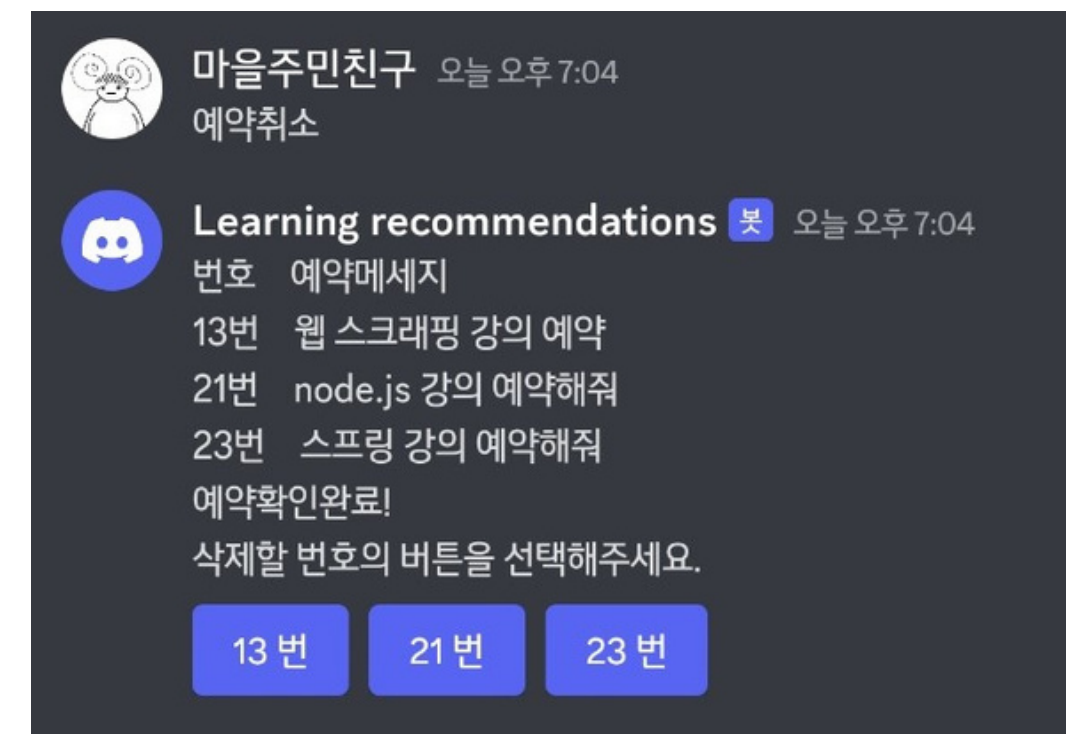
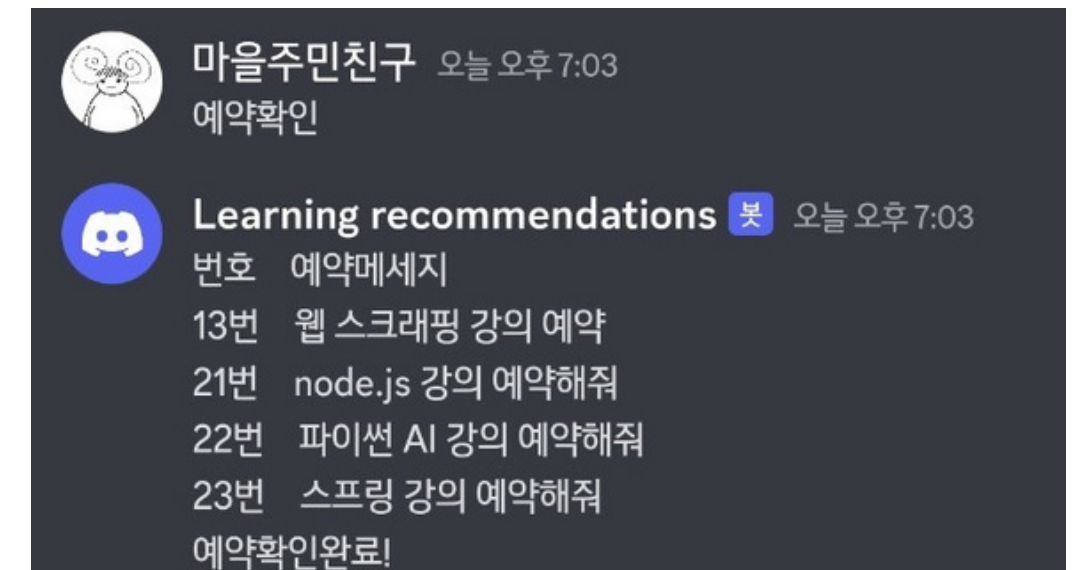
실제 화면



강의 예약 키워드로 강의 예약



예약 강의 업로드 알림
예약 키워드와 관련된 새 강의가
업로드 되었을 때 메시지로 알려줌



예약 확인 및 취소
예약했던 키워드를 확인하고 예약 취소도 가능

기술셋

기본 언어



자료 수집



BeautifulSoup



추천 모델



백엔드



추후 업데이트 사항 - 데이터 수집

1. 새롭게 추가되는 강의들 크롤링 자동화 작업
2. 보기쉽게 코드 간결화
3. 다른 인터넷 강의 사이트 크롤링 및 전처리

추후 업데이트 사항 - 추천 모델

1. 랭킹에 들어가는 항목 가중치 조정

2. 더 많은 유저 로그가 쌓인 후 추천 개인화 - 협업 필터링 등

추후 업데이트 사항 - 백엔드

1. 사전이 있어 사용자가 넣을수 있는 모든 경우의 수를 db에 적재하여 속도개선
2. 디스코드 봇 슬래쉬 커맨드를 사용하여 편의성 개선

트러블슈팅 및 배운점

백엔드

처음 구름에 환경을 구축하려고 할때 디스코드 봇 라이브러리 설치가 안되어 찾아보니 구름에서 제공하는 제일 높은 버전보다 높은 버전인 3.8이상 에서만 동작하여 버전을 업그레이드 하였고 pip또한 버전문제로 새로 업그레이드 하였고 인터넷에 올라온 예제들의 버전이 다 달라서 원하는 내용을 찾는것이 힘들었지만 이러한 일을 하게되면서 필요한 내용을 찾을때 놓아야하는 키워드를 좀더 잘 찾을 수 있게되었습니다.

추천 모델

Doc2Vec 모델을 사용하기 위해 Gensim이라는 라이브러리를 사용해야 했는데 제가 공부한 코드가 에러를 일으켜서 당황했는데 구글링 해보니 이 라이브러리가 최근 업데이트를 하면서 파라미터 입력 방식을 대거 바꾼 관계로 이전의 코드를 실행하면 오류가 났던 것이었습니다. 그래서 최근 문서를 보면서 파라미터를 직접 수정했는데 이를 통해 작업할 때 필요한 라이브러리 버전을 맞추는게 중요함을 다시 한번 깨달았습니다.

데이터 수집

강의 리뷰를 get방식으로 가져오려고 했는데 요청url에 필요한 강의id가 파싱해서 사용하고 있던 html에 없어서 찾는데 시간이 오래 걸렸던 문제가 있었습니다. 셀레니움으로 버튼을 눌러가며 모든 리뷰를 모을 수는 있지만 그렇게 되면 시간이 오래 걸리기에 강의id를 어디서 가져올 수 있을지 생각해보고 홈페이지를 분석해 갔습니다. 그리고 강의 하나하나의 html이 아닌 모든 강의가 보이는 메인의 html에서 강의id를 가지고 올 수 있음을 알았고 무사히 get방식으로 데이터를 수집할 수 있었습니다. 다른 데이터들은 각 강의의 html에서 데이터를 가지고 왔기에 리뷰도 그렇게 해야한다고 저도 모르게 시야가 좁아져 있었지만 넓게 보니 해결방법이 보였던 이번 경험을 통해 찾는게 안보인다고 생각되는 상황에서도 흐름을 이해하고 시야를 넓혀 크게 보면 찾을 수 있다는 것을 배울 수 있었습니다.

후기

영현

팀원들의 개발 코드를 공유하는 과정에서 처음 보는 사람도 읽기 좋고 이해할 수 있는 코드를 짜는것의 중요성을 알게되었습니다. 그리고 이번 부트캠프에서 배웠던 것을 다시한번 복습하고, 정리할 수 있었던 시간이었던 것 같습니다.

민호

팀원 분들과 함께 개발하면서 코드를 공유 할때 서로가 사용할때 입력 및 출력을 제대로 정하지 않아서 서로의 코드를 연결할때 추가로 일을 하게되어 처음에 멘토님이 말씀하셨던 요구사항명의를서가 자세할수록 좋다는 말씀이 더욱 와 닿았습니다. 그리고 개발을 하면서 원하는기능이 제대로 작동하게 하였을때 성취감이 있어 새로운 기능을 추가하는게 즐거웠던거 같습니다.

소영

추천 알고리즘에 대해 알아보면서 웹서비스의 많은 부분이 알고보면 추천 시스템임을 배웠습니다. 또한 팀프로젝트 멘토링 경험이 풍부하신 멘토님께서 프로젝트의 기획단계부터 마지막 발표까지 꼼꼼히 체크하고 조언해주셔서 많은 도움을 받았습니다.

감사합니다

조소영 - sojo6658@gmail.com
고영현 - sns5877@naver.com
양민호 - kinding45@gmail.com