

비인기도서 대출 활성화를 위한 생성형 AI도서추천 챗봇 및 메일서비스

Alpaco AI엔지니어 부트캠프 5기

5조: 민경환 김철 유태호 최종현

목 차

- 01. 프로젝트 개요
- 02. 프로젝트 구성 및 역할
- 03. 프로젝트 수행 절차 및 방법
- 04. 프로젝트 추진 배경
- 05. 프로젝트 수행 결과
- 06. 자체 평가 및 의견

01. 프로젝트 개요

1. 프로젝트 명 : 비인기도서 대출 활성화를 위한 생성형 AI도서추천 챗봇 및 메일 서비스
2. 프로젝트 개발기간 : 2023. 6. 16. ~ 2023. 6. 26.
3. 활용 공공데이터 : 문화 빅데이터 <https://www.bigdata-culture.kr/bigdata/user/main.do>
도서별 상세정보 / 도서별 키워드 정보 / 도서별 대출추이분석 /
도서별 이용정보분석 / 도서별 대출 반감기 (국립중앙도서관)
4. 수행 프로그램 : Python, Python Dash, UIPATH
5. 개발담당자 : 민경환, 김철, 유태호, 최종현

02. 프로젝트 구성 및 역할

훈련생	역할	담당 업무
민경환	팀장	아이디어 개발, 데이터 수집, 데이터 전처리, 구독서비스 개발(주), RPA 메일발송 서비스 개발(주)
김 철	팀원	아이디어 개발, 데이터 수집, 데이터 전처리, 챗봇서비스 개발(주), Dash 디자인(주)
유태호	팀원	아이디어 개발, 데이터 수집, 데이터 전처리, 구독서비스 개발(부), RPA 메일발송 서비스 개발(부)
최종현	팀원	아이디어 개발, 데이터 수집, 데이터 전처리, 챗봇서비스 개발(부), Dash 디자인(부)

03. 프로젝트 수행 절차 및 방법

구분	기간	활동	비고
사전 기획	▶ 6/16(금), 6/19(월)	▶ 프로젝트 아이디어 수집 및 유효성 검토 ▶ 프로젝트 목표 수립 및 DB소재처 검색	
데이터 수집	▶ 6/19(월) ~ 6/20(화)	▶ 데이터 포털에서 관련 자료 전체 수집 및 선별	▶ 문화 빅데이터 플랫폼 ▶ 도서관 정보나루
데이터 전처리 및 분석	▶ 6/20(화) ~ 6/22(목)	▶ 선별데이터 클리닝 ▶ 데이터 시각화 및 분석(Python) ▶ 활용데이터 베이스 구축	▶ Python
서비스 구축	▶ 6/22(목) ~ 6/23(금)	▶ ChatGPT와 Dash를 이용한 챗봇 서비스 구현 ▶ Dash와 RPA를 이용한 구독 및 메일 서비스 구현	▶ Python Dash ▶ UIPATH
서비스 구축 및 보수	▶ 6/23(금) ~ 6/26(월)	▶ 통합서비스 웹페이지 및 RPA 구축, 배포	▶ 최적화 및 오류 수정
총 개발기간	▶ 6/16(화) ~ 6/26(월)	-	-

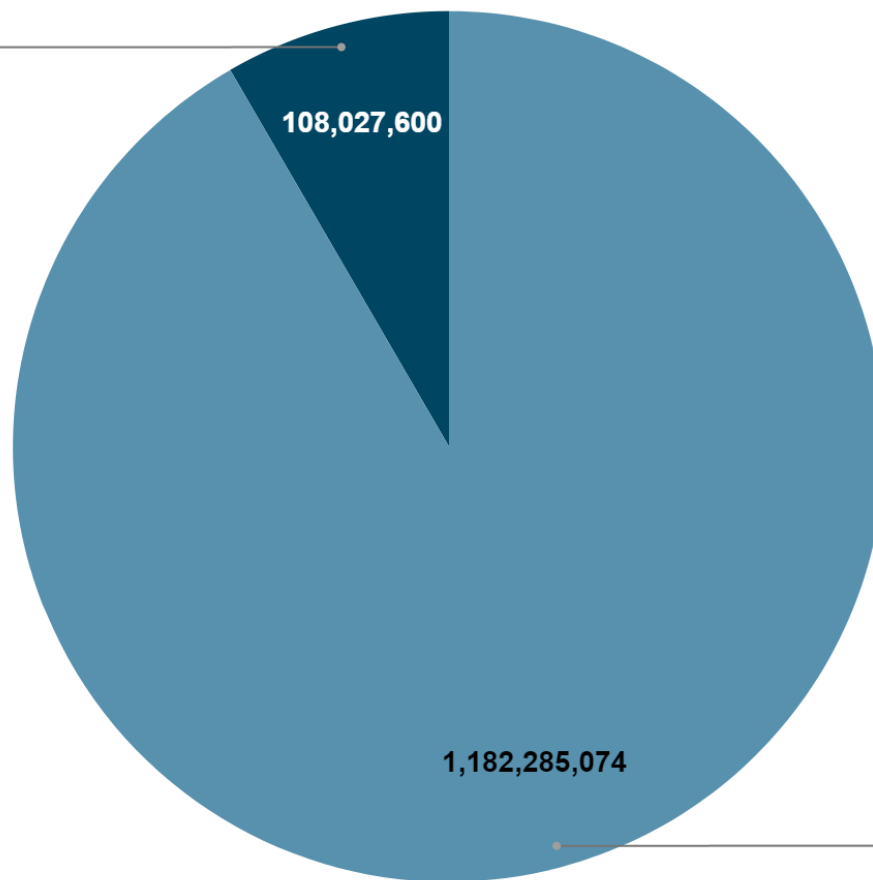
04. 프로젝트 추진 배경

- 좋은 도서관은 책장이 비어있는 도서관이다.
- 서가에 꽂혀 있는 다양한 책들을
시민들이 읽을 수 있도록 데이터를 활용하려면?

04. 프로젝트 추진 배경

2021 도서 이용정보 분석: 대출 상위 25%, 대출 하위 75% 대출수 합 비교

하위 75%(817,970권)
8.4%



상위 25%(272,785권)
91.6%

상위 25% 책들의 대여수 총합은 하위 75%의 약 10배
→ 인기 도서에만 쏠리는 대중의 관심을 비인기 도서로 분산시킬 필요성이 있음.

05. 프로젝트 수행결과

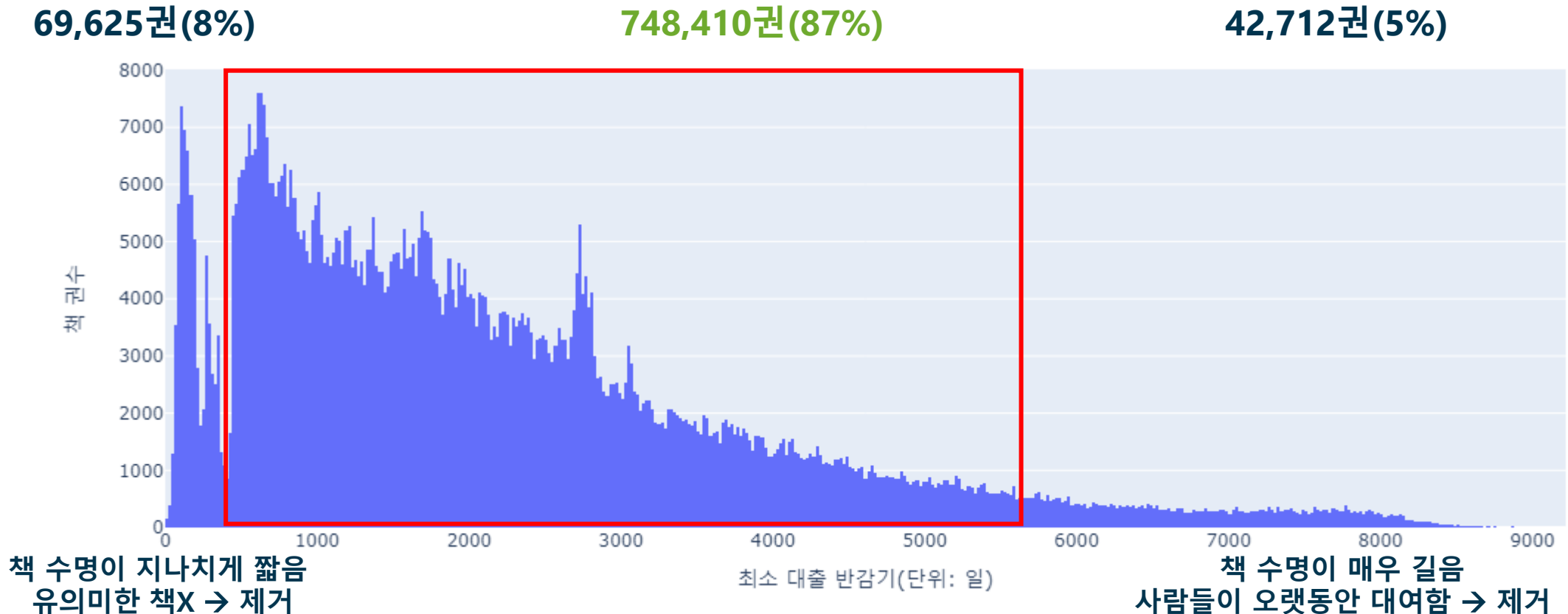
1. 수집 데이터 분석 및 전처리

- 데이터 병합(Merge) 및 중복 컬럼 제거
 - 도서별 상세정보 / 도서별 키워드 정보 / 도서별 대출추이분석 /
도서별 이용정보분석 / 도서별 대출 반감기 (국립중앙도서관)
- 병합시 기준 데이터테이블 : 도서별 대출 반감기 (약 110만건)
- 110만권 중 KDC 분류 기호(도서별 상세정보)가 있는 데이터 860,747건 추출
- 키워드(검색어) 정보(약 13만건)는 검색에서 부수적으로 활용

05. 프로젝트 수행결과

1. 수집 데이터 분석 및 전처리

- 기준데이터 선정: 도서별 대출 반감기(국립중앙도서관) → 책 수명에 관한 정보
- 사람들이 적당히 흥미있게 빌렸으나, 책 수명이 길지도, 짧지도 않은 구간을 선별, 추출



05. 프로젝트 수행결과

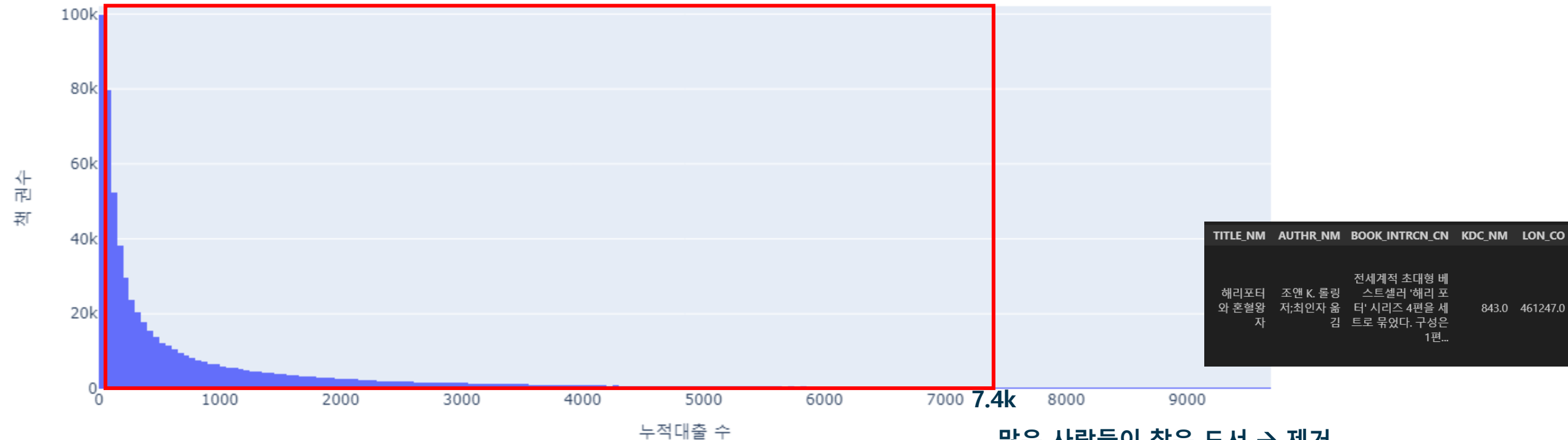
1. 수집 데이터 분석 및 전처리

- 기준데이터 선정: **도서별 이용정보 분석**(국립중앙도서관)
- 대여수가 상하위 5% 구간 제거 (단, 결측치 제외하였으며 결측치는 평균값으로 대체)

35,390권(5%)

662,455권(95%)

35,075권(5%)



많은 사람들이 찾은 도서 → 제거

05. 프로젝트 수행결과

2. 전처리 데이터 가공

- 책 검색에 필요한 정보 취합
 - 키워드(검색어), 서지 정보 및 책 소개, KDC 분류기호

- 전체 데이터에서 Chatbot이 사용할 DB파일 추출 (최종 700,407권)

```
df.columns
✓ 0.0s

Index(['BOOK_MASTR_SEQ_NO', 'ISBN_NO', 'MUMM_LON_HALFLIFE_CO',
      'MXMM_LON_HALFLIFE_CO', 'AVRG_LON_HALFLIFE_CO', 'BOOK_CO', 'SEQ_NO',
      'ISBN_THIRTEEN_NO', 'TITLE_NM', 'AUTHR_NM', 'PUBLISHER_NM', 'IMAGE_URL',
      'BOOK_INTRCN_CN', 'KDC_NM', 'LON_CO', 'Combined', 'KDC_1', 'KDC_12',
      'KDC_123'],
      dtype='object')
```

```
db.sample(5)
✓ 0.0s
```

	ISBN_NO	TITLE_NM	AUTHR_NM	PUBLISHER_NM	BOOK_INTRCN_CN	Combined	KDC_1
738221	9788901214498	인지니어들:실리콘밸리 인재의 산실 '스탠퍼드 디스쿨'의 기상천외한 창의력 프로젝트	티나 실리그 지음;김소희 옮김	웅진씽크빅	'디스쿨'을 직접 이끄는 교수진 중 한 명인 티나 실리그 교수가 밝히는 스탠퍼드 ...	매출액, 타입, 유수, 전면, 철자, 평상복, 메트, 굽다, 실용적, 버튼, 경쟁...	1
90028	9788936485283	미국의 엔진, 전쟁과 시장	김동준 지음	창비	사회학자 김동준 교수의 미국 깊이 읽기. 자본주의 세계체제, 보수기독교, 이라크 전...	교착, 원장, 단재, 세기, 중심, 제국경영, 이라크, 시민, 선별, 폐해, 시장근...	3
315767	9788933108437	나무야, 너는 아니?=My tree, do you know?	최용진 글,한용욱 그림	바오로딸	지난 겨울 빼만 남아 앙상했던 은행나무가 봄이 되자 다시 아기 울음을 터트리며 깨어...	NaN	8
211923	9788985320689	(이원호 장편소설)밤의 대통령	이원호 지음.	문학수첩	이 소설이 3백만 독자를 사로잡은 까닭은 우리 모두가 기다렸던 영웅을 탄생시켰기 때...	NaN	8
752280	9791125587286	파병부대 =이윤규 장편소설 /Dispatch troops	지은이: 이윤규	Rok Media(로크미디어)	이윤규 장편소설. 2030년, 평화 유지를 위해 출발한 남북한 통합 파병 부대. 그...	NaN	8

05. 프로젝트 수행결과

3.1. 도서추천 ChatBot 서비스 개발

- ChatGPT API(모델: "text-davinci-003") 사용
- ChatGPT API를 이용하여 Chatbot 서비스에서 발생할 수 있는 Hallucination 문제
→ 활용 DB안에서 답을 제공 할 수 있도록 Query조건 강화
- Dash에서 채팅UI 적용

```
final_question = """
책을 추천하는 사서라고 생각하고 말을 해야해
아래 데이터프레임을 파싱해서
데이터 프레임 column은 다음과 같이 파싱할 수 있다
1. 데이터프레임 헤더 정보: {}
2. 데이터프레임: {}

책 제목, 저자, 출판사, ISBN을 알려주고
간단한 책 내용을 요약해서 알려준다
알려줄 정보는 데이터프레임에 있는 것만 알려줄 것
가독성이 좋도록 줄바꿈을 넣어 줄 것
""".format(column_description, query_result)
```

05. 프로젝트 수행결과

3.1. ChatGPT API 활용하여 DB쿼리, 책 추천 서비스 개발

- 키워드 검색

도서추천 봇

도서추천 메일서비스

안녕하세요
도서추천 봇입니다.

테니스 책 추천해줘

베스루테니스: 저자 Alan S. Boltin, 출판사 홍경. 테니스 교본. 베스루테니스의 개념 소개를 시작으로 볼과 친해지는 법과 샷하기 전까지의 기본적인 기술을 소개한다. ISBN: 9788984441095

(알고보면 쉬운)멘탈 트레이닝:테니스: 저자 김병준, 출판사 무지개사. 이 책은 저자가 테니스 코리아의 예 3년동안 연재한 내용을 재구성한 것이다. 1,2년 전부터 시작해서 최근까지의 멘탈 트레이닝 방법을 소개한다. ISBN: 9788989377276

최신 테니스 지도의 이론과 실제: 저자 이충훈, 출판사 경남대학교출판부. 테니스 지도의 이론과 실제를 다룬 체계적인 지도서. 저자가 선수생활 시절부터 20여년간 지도해온 경험을 바탕으로 테니스 지도자를 위한 이론과 실제를 소개한다. ISBN: 9788984210868

테니스 책 추천해줘

질문하기

- 맥락, 상황 검색

도서추천 봇

도서추천 메일서비스

는 세계를 만들어갈 수 있는지 보여주는 책.
소년 로난이 작은 풀장에서 발견한 물 속 세계를 방문하며 다양한 모험을 겪는 이야기.

(지구를 굴리는)식종구리 (안선모 지음;전가영 그림, 문공사 출판사, ISBN 9788945214041)
공부를 특별히 잘 하는 것도 아니고, 무언가를 뛰어나게 잘 하는 것도 아니고, 그저 일반적인 소년 식종구리가 가진 가치를 발견하는 이야기.
식종구리가 지구를 굴리며 자신의 가치를 발견하는 이야기.

교양을 키울 수 있는 책 추천해줘

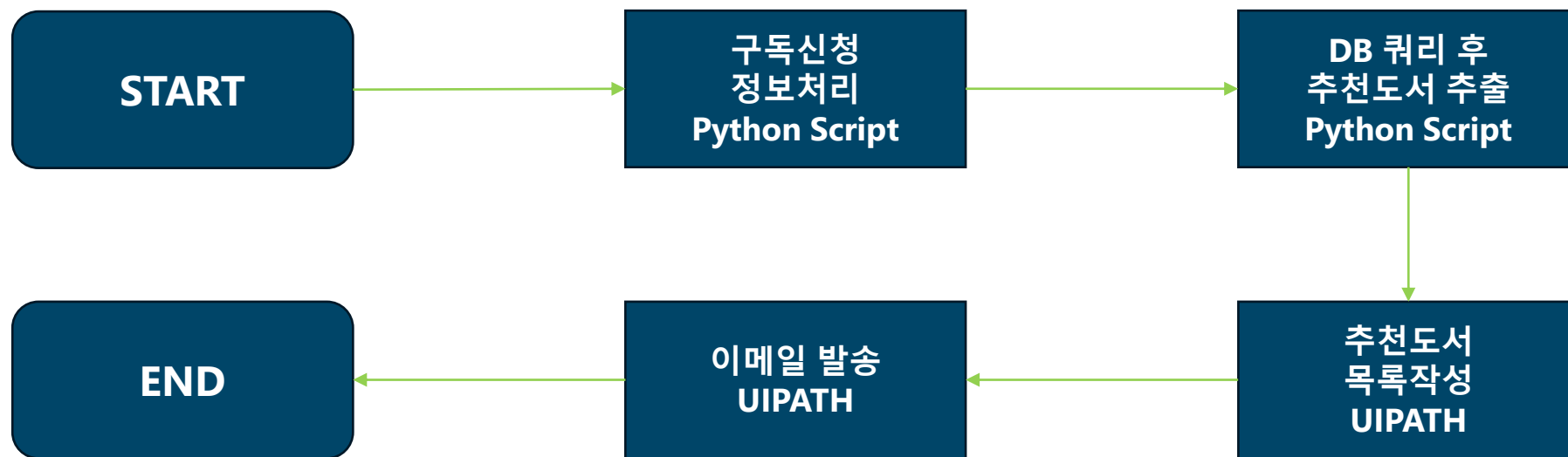
오늘 추천할 책은 '경제가 쉽다 경제기사가 즐겁다'입니다. 장동학 지음, 무한 출판사에서 발행된 책입니다. ISBN 번호는 9788988800737 입니다. 일반인들이 쉽게 이해할 수 있도록 풀어쓴 경제교양서로, 경제 성장·국민소득, 경기, 물가, 금융, 외환 등 경제에 관한 기본적인 이론과 실제를 쉽게 이해할 수 있도록 설명하고 있습니다.

교양을 키울 수 있는 책 추천해줘

질문하기

05. 프로젝트 수행결과

3.2. UIPATH(RPA)를 활용한 EMAIL 책 추천 서비스



3.2. UIPATH(RPA)를 활용한 EMAIL 책 추천 서비스

RPA로 Python 스크립트 실행

도서추천 봇

도서추천 메일서비스

양자역학

자연과학

구독하기

구독을 신청하였습니다.



	A	B	C	D
1	EMAIL	KDC	SRCHWRD	DATE
2			4 양자역학	2023-06-23
3			4 ㄱ ㄴ ㄷ ㄹ	2023-06-23
4			4 양자역학	2023-06-23
5			4 양자역학	2023-06-23

EMAIL 중복시 최후 신청건만 활용

	A	B	C	D	E	F	G	H
1	ISBN_NO	TITLE_NM	AUTHR_NM	PUBLISHER_NM	IMAGE_URL	BOOK_INTRC	KDC_NM	Email_recommend
2	9.79E+12	양자역학 =Quantum mechanics	저자: 최준곤	범한서적주식회사	http://image.aladin.co.kr/product/763/30/cover/8971292326_1.jpg	『양자역학』은 양자역학의 필요성에 부응하여 양자역학을 공부하는 학생들에게 도움이 되는 내용을 제공한다. 양자역학의 기본적 개념에 대한 인	420.13	책 이름: 양자역학 =Quantum mechanics 책 저자: 저자 최준곤 출판사: 범한서적주식회사 책 소개: 『양자역학』은 양자역학의 필요성에 부응하여 양자역학을 공부하는 학생들에게 도움이 되는 내용을 제공한다. 양자역학의 기본적인 내용을 참고하고, 연구에 적용할 수 있는 여러 응용 분야에 대한 설명과 이에 덧붙여 심도 있는 문제들을 연습문제에서 풀어볼 수 있게 구성하였다. KDC분류기호: 420.13
3	9.79E+12	양자역학의 철학적 기초	한스 라이헨바흐 지음 ;강형구 옮김	지식을만드는지식	http://image.aladin.co.kr/product/3898/35/covers/1130412059_1.jpg	계에 대한 인간의 지식에 어떤 함축을 갖는가? 이 책은 유럽에서 ‘논리경험주의’를 이끌었던 과학철학자 한스 라이헨바흐가 양자역학	420.13	ISBN: 9791130412054 책 이름: 양자역학의 철학적 기초 책 저자: 한스 라이헨바흐 지음 ;강형구 옮김 출판사: 지식을만드는지식 책 소개: 양자역학은 세계에 대한 인간의 지식에 어떤 함축을 갖는가? 이 책은 유럽에서 ‘논리경험주의’를 이끌었던 과학철학자 한스 라이헨바흐가 양자역학을 철학적으로 분석한 저술이다. 양자역학은 고전역학과 어떤 차이가 있는지, 양자역학의 언어에 적합한 논리는 무엇인지를 살핀다. KDC분류기호: 420.13

🔍 검색 → 예외상황 대응

	A	B	C	D	E	F	G	H
1	ISBN_NO	TITLE_NM	AUTHR_NM	PUBLISHER_NM	IMAGE_URL	BOOK_INTRC	KDC_NM	Email_recommend
2	9.79E+12	(만화) 사이언스 21:식물	기획: 신나는 과학을 만드는 사람들;만화: 박종호	삼성출판사	https://bookthumb-phinf.pstatic.net/cover/030/146/03014662.jpg?type=m1&date=20110810	현직 초중고 과학 선생님이 기획 · 집필한 과학만화! 『만화 사이언스 21』 시리즈 『식물』. 본 시리즈는 현직 초중고 과학 선생	08	책 이름: (만화) 사이언스 21:식물 책 저자: 기획 신나는 과학을 만드는 사람들;만화 박종호 출판사: 삼성출판사 책 소개: 현직 초중고 과학 선생님이 기획 · 집필한 과학만화! 『만화 사이언스 21』 시리즈 『식물』. 본 시리즈는 현직 초중고 과학 선생님들로 구성된 '신나는 과학을 만드는 사람들'이 기획 · 집필한 과학 교양서로, 어린이의 눈높이에 맞게 만화로 풀어냈습니다. 특히 만화 속에 최신 과학 정보를 흥미롭게... KDC분류기호: 408
3	9.79E+12	일반화학 탐구실험 :창의적 문제 해결력 지향	강순희,우애자	북스힐	http://image.aladin.co.kr/product/248/41/covers/8955265131_1.jpg	일반화학 교재. 화학전공자, 화학관련 실무자들을 위해 일반화학의 이론을 설명하였다.	32	ISBN: 9788955265132 책 이름: 일반화학 탐구실험 :창의적 문제 해결력 지향 책 저자: 강순희,우애자 출판사: 북스힐 책 소개: 일반화학 교재. 화학전공자, 화학관련 실무자들을 위해 일반화학의 이론을 설명하였다. KDC분류기호: 432

05. 프로젝트 수행결과

3.2. UIPATH(RPA)를 활용한 EMAIL 책 추천 서비스

☆ [책 추천] 이런 책 어떠세요? 📧

📧 인쇄 | 번역

^ 보낸사람 전자도서관사서 [redacted] VIP

받는사람 [redacted]

2023년 6월 23일 (금) 오후 1:14

안녕하세요.

오늘도 많은 책들이 도서관에서 나오고 싶어합니다.

오늘 도서관 탈출을 꿈꾸는 후보 5권을 추천해드립니다.

지식 가득한 하루 되시길 바랍니다!

ISBN: 9788935801565

책 이름: 테니스 바이블

책 저자: 朴勝和 著

출판사: 保景文化社

책 소개: <https://www.aladin.co.kr/search/wsearchresult.aspx?SearchTarget=All&SearchWord=9788935801565>

KDC분류기호: 695.5

ISBN: 9788990784681

책 이름: 테니스가 즐거워지는 스트레칭과 트레이닝 =Stretching & training

책 저자: 저자 조성연,테니스코리아 편집부

출판사: 미디어월 월간 테니스코리아

책 소개: 스포츠 활동을 더욱 안전하고 효과적으로! 『테니스가 즐거워지는 스트레칭과 트레이닝』. 이 책은 테니스 관련 전문가는 물론 일반인도 쉽게 실천할 수 있는 다양한 체력 트레이닝에 관한 내용을 담았다. 부상 없이 즐겁게 테니스를 즐길 수 있는 스트레칭 방법과 트레이닝 기술 22가지를 소개했다.본서의...

KDC분류기호: 695.5

책소개가 없는 경우, 인터넷 서점 도서ISBN 검색링크 제공

☆ [책 추천] 이런 책 어떠세요? 📧

📧 인쇄 | 번역

^ 보낸사람 전자도서관사서 [redacted] VIP

받는사람 [redacted]

2023년 6월 23일 (금) 오후 11:35

안녕하세요.

오늘도 많은 책들이 도서관에서 나오고 싶어합니다.

오늘 도서관 탈출을 꿈꾸는 후보 5권을 추천해드립니다.

지식 가득한 하루 되시길 바랍니다!

ISBN: 9788971292327

책 이름: 양자역학 =Quantum mechanics

책 저자: 저자 최준곤

출판사: 범한서적주식회사

책 소개: 『양자역학』은 양자역학의 필요성에 부응하여 양자역학을 공부하는 학생들에게 도움이 되는 내용을 제공한 다. 양자역학의 기본적인 내용을 참고하고, 연구에 적용할 수 있는 여러 응용 분야에 대한 설명과 이에 덧붙여 심도 있는 문제들을 연습문제에서 풀어볼 수 있게 구성하였다.

KDC분류기호: 420.13

ISBN: 9791130412054

책 이름: 양자역학의 철학적 기초

책 저자: 한스 라이헨바흐 지음 ;강형구 옮김

출판사: 지식을만드는지식

책 소개: 양자역학은 세계에 대한 인간의 지식에 어떤 함축을 갖는가? 이 책은 유럽에서 '논리경험주의'를 이끌었던 과학철학자 한스 라이헨바흐가 양자역학을 철학적으로 분석한 저술이다. 양자역학은 고전역학과 어떤 차이가 있는 지, 양자역학의 언어에 적합한 논리는 무엇인지를 살핀다.

KDC분류기호: 420.13

↑

06. 자체 평가 및 의견

• 장점

- 도서관에서 잠들어있는 책들을 추천할 수 있기 때문에, 도서관 이용률을 더욱 높일 수 있음.
- 대쉬보드 서비스를 UIPATH를 이용하여 서비스 구동 및 배포과정을 간편하게 할 수 있도록 하였고 UIPATH를 이용하여 Python Script, Excel, Send Mail 과정을 한번에 처리한 점.
- 중소형 도서관 시스템 DB활용하여 바로 적용가능 → 중소형 홍보 및 활성화에 기여할 수 있음.
- DB라벨링 및 NLP 인코딩할 필요 없이, GPT모델이 DB에 해당하는 무난한 정확도의 답을 도출시키도록 한 점. (Hallucination 방지)
- 현재의 책 추천서비스는 인기도서, 사서추천 도서, 광고도서에 한정되어 있는데, 다양한 종류의 책을 추천할 수 있음.

• 개선할 점

- 추천 도서를 도서관 대여 서비스와 연계할 필요성이 있음 (추후 개발 과제)
- 추가 자료를 수집하여 자연어 처리를 통해 도서정보, 키워드들을 연결시켜 단순 쿼리를 넘어서 유사 도서를 추천할 수 있도록 개선 (추후 개발 과제, 라벨링 문제 해결 필요)
- KDC 분류번호에 대한 자료 정보가 제한적인 점 (DB개선 및 요청 필요)
- ChatGPT에 질의할 때 발생하는 예상치 못한 답변을 줄일 필요성이 있음. (쿼리조건 개선 필요)

Q&A