

PJT 02

Python을 사용한 수집 데이터 활용

• INDEX

- Aladin API 데이터 활용
 - 준비 사항
 - 요구 사항
 - 과제 그룹 A, B, C
 - 도전 과제
- 제출 안내

프로젝트 개요

- Aladin 도서 데이터를 직접 수집하고 이를 활용하는 단계로
주어진 데이터를 분류하고 응용하는 과정

프로젝트 목표

- Python 기본 문법 습득
- 파일 입출력에 대한 이해
- 데이터 구조에 대한 분석과 이해
- 텍스트 - 음성 변환 (TTS)
- 생성형 AI를 활용한 데이터 분류

Aladin API 데이터 활용

준비 사항

개발 도구 및 라이브러리

- 개발 도구
 - Visual Studio Code
 - Python 3.9+
 - requests
 - <https://docs.python-requests.org/en/latest/index.html>
 - pathlib
 - <https://docs.python.org/3.9/library/pathlib.html>
 - gTTS
 - <https://gtts.readthedocs.io/en/latest/>
 - json
 - <https://docs.python.org/3.9/library/json.html>

제공 사항

- 기본 스켈레톤 파일

요구 사항

과제 그룹 별 요구 사항

- 과제 그룹 A
 - API 요청 함수 작성하기
 - 도서 데이터 가공 및 저장
 - TTS 파일 생성하기
- 과제 그룹 B
 - 다양한 주제 데이터 수집 및 통합하기
 - 데이터 분류 및 저장하기
 - TTS 활용하여 책 데이터 요약 오디오 만들기

과제 그룹 별 요구 사항

- 과제 그룹 C
 - 데이터 검색 기능 확장 및 분석하기
 - 생성형 AI를 활용한 키워드 기반 도서 분류하기
 - TTS를 활용하여 주제별 오디오 파일 생성하기
- 도전 과제
 - 도서 데이터를 생성형 AI를 활용하여 요약하고 이를 오디오로 생성해보기
 - 챗봇 동작을 할 수 있도록 전체 코드를 정리해보기

과제 그룹 A

API 요청 함수 작성하기

목표: fetch_books 함수를 완성하여 검색된 도서 정보를 출력합니다.

조건:

1. API 키는 .env 파일에 저장하며, `dotenv` 라이브러리를 통해 불러옵니다.
2. `requests` 라이브러리를 활용하여 검색 요청을 보냅니다.
3. 응답 데이터를 JSON 형식으로 변환 후 fetch_books 함수의 실행 결과로 반환합니다.
4. 10개의 " 펭귄 " 관련 도서를 넘버링과 함께 제목을 출력합니다.

• 결과 화면

1. 펭귄도 모르는 남극 이야기
2. 너에게 들려주는 단단한 말 (윈터 패키지 에디션, 특별 한정 필사노트 포함)
3. 남극 펭귄 생포 작전
4. 핑크펭귄 - 완전히 새롭고 전혀 다른 그 무언가!, 개정판
5. 펭귄 365
6. 나를 소모하지 않는 현명한 태도에 관하여
7. 최재천의 동물대탐험 5 : 황제펭귄의 행진 - 남극의 육아 천재, 황제펭귄
8. 매거진 B (Magazine B) Vol.10 : 펭귄박스 (PENGUIN) - 국문판 2012.10
9. 펭귄 호텔
10. 펭귄은 너무해

도서 데이터 가공 및 저장

목표: 특정 키워드("컴퓨터")로 검색된 도서 정보를 저장하고 필요한 필드만 정리하여 파일로 저장 합니다.

조건:

1. 검색 결과를 반환하는 fetch_books 함수를 완성합니다.
2. 파일 저장을 위한 'output' 폴더를 생성합니다. (이미 폴더가 존재해도 에러를 발생시키지 않도록 합니다.)
3. save_books_to_file 함수는 검색 결과 데이터를 받아와 지정된 파일에 저장합니다.

3-1. 각 도서는 "제목", "저자", "출판사", "가격" 항목으로 나누며 마크다운 형식으로 출력 및 파일로 저장 합니다.
4. 도서 데이터는 output/computer_books.txt 파일로 저장됩니다

• 결과 화면

```
## 제목: 지금 당장 양자컴퓨터에 투자하라 - AI를 넘어설 퀀텀 혁명,  
* 저자: 권중현(가을바람) (지은이)  
* 출판사: 애덤스미스, * 가격: 24000원  
-----  
## 제목: 2025 이기적 컴퓨터그래픽기능사 필기 절대죽보 - 변경 출제기준 1  
00% 반영 + CBT 온라인 문제집 + 동영상 강의 무료 제공,  
* 저자: 김부치 (지은이)  
* 출판사: 영진.com(영진닷컴), * 가격: 20000원  
-----  
## 제목: 2025 이기적 컴퓨터그래픽기능사 필기 기보서 - 변경 출제기준 100
```

(중략)

```
## 제목: 2025 해커스공무원 각후근 컴퓨터일반 단원별 기출문제집 (전산직  
공무원) - 전산직, 계리직, 군무원, 소방경채 시험 대비 | 컴퓨터일반 무료  
특강 | 합격예측 온라인 모의고사 응시권 제공,  
* 저자: 각후근 (지은이)  
* 출판사: 해커스공무원, * 가격: 27000원  
-----
```

(실시간 데이터이기에 출력 결과가 상이할 수 있음)

TTS 파일 생성하기

목표: `gTTS` 라이브러리를 이용해 검색된 도서의 첫 번째 제목을 음성 파일로 저장합니다.

조건:

1. “fetch_book_title” 함수는 주어진 키워드를 사용하여 알라딘 API에서 도서를 검색하고 제목을 하나 반환합니다..
2. “create_tts_file” 함수는 “fetch_book_title”을 호출하여 도서 제목을 가져오고, 해당 제목을 텍스트 음성 변환(TTS)하여 오디오 파일로 저장합니다.
3. 파일 경로는 “output/book_title.mp3”로 저장합니다.

• 결과 화면

음성 파일이 output/book_title.mp3로 저장되었습니다.

▼ output
🔊 book_title.mp3

과제 그룹 B

다양한 주제 데이터 수집 및 통합하기

목표: '자연', '예술', '기술' 세 가지 주제에 대해 각각 30개의 도서를 조회하고 평균 가격과 평균 출판 연도를 출력합니다.

조건:

1. 주제별로 도서 정보를 조회하여 수집하는 `fetch_books_by_topic` 함수를 완성합니다.
2. `topic_stats` 딕셔너리에 각 주제별 통계 (평균 가격, 평균 출판 연도)를 계산하여 저장합니다.
3. 모든 책의 평균 가격과 평균 출판 연도를 계산하여 저장 합니다.
4. 결과를 화면에 출력합니다.

• 결과 화면

```
통합된 도서 수: 90
전체 통계: 평균 가격 21344원, 평균 출판 연도 2020
자연 주제 통계:
  평균 가격: 22100.00원
  평균 출판 연도: 2019
예술 주제 통계:
  평균 가격: 21893.33원
  평균 출판 연도: 2020
기술 주제 통계:
  평균 가격: 20040.00원
  평균 출판 연도: 2022
```

(실시간 데이터이기에 출력 결과가 상이할 수 있음)

데이터 분류 및 저장하기

목표: '경제' 관련 도서 데이터를 50개 수집하고, 할인 중인 책과 정가인 책으로 나누어 JSON 형식으로 저장합니다.

조건:

1. 주제별로 도서 정보를 조회하여 수집하는 `fetch_books_by_topic` 함수를 완성합니다.
2. 할인 가격(`priceSales`)와 정가(`priceStandard`)를 비교하여 할인 여부를 판단합니다.
3. 할인 중인 책 데이터는 ``discounted_books.json``, 정가인 책 데이터는 ``regular_priced_books.json`` 파일에 저장합니다.
4. 할인 중인 책의 개수를 출력합니다.

• 결과 화면

할인 중인 책 2개, 정가인 책 48개를 저장했습니다.

(실시간 데이터이기에 출력 결과가 상이할 수 있음)

TTS를 활용하여 데이터 요약 오디오 만들기

목표: "한국문학" 주제를 검색하여 description이 있는 5개의 도서 제목과 요약을 JSON으로 정리합니다.

조건:

1. 주제별로 도서 정보를 조회하여 수집하는 `fetch_books_by_topic` 함수를 완성합니다.
2. “`save_books_to_json`” 함수는 도서에 'description' (요약) 정보가 있을 경우에만 리스트에 총 5개를 담고, 이를 “`output/korean_literature_books.json`” 파일로 저장 합니다.
3. “`create_tts_file`” 함수는 저장된 JSON파일을 읽어 한국어 음성으로 변환하고, “`output/book_n.mp3`” 파일로 저장합니다.
4. 각 도서의 제목과 요약을 텍스트로 결합하여 각각의 오디오 파일(`book_n.mp3`)을 생성합니다.

TTS를 활용하여 데이터 요약 오디오 만들기

결과 화면

5개의 책 정보를 처리하고 음성 파일로 저장했습니다.

```
▼ output
🔊 book_1.mp3
🔊 book_2.mp3
🔊 book_3.mp3
🔊 book_4.mp3
🔊 book_5.mp3
📄 korean_literature_books.json
```

과제 그룹 C

데이터 검색 기능 확장 및 분석하기

목표: 인공지능' 관련 도서 100개를 알라딘 API에서 검색하고, 가격이 높은 순서대로 상위 10개의 도서를 JSON 파일로 저장합니다.

조건:

1. 알라딘 API는 한 번에 50개의 데이터만 응답으로 옵니다. 문서를 참고하여 중복되지 않은 100개의 책의 데이터를 수집하는 함수 `fetch_book_by_topic` 함수를 완성 합니다.
2. 100개의 도서 데이터 중 가격이 높은 순서대로 상위 10개의 도서 정보를 출력합니다.
3. 제목과 저자 정보를 요약하여 `'ai_top10_books.json'` 파일로 저장합니다.

• 결과 화면

가격이 높은 순서대로 상위 10개 도서:

1. 제목: the AI GRAPHICS : 인공지능, 캐릭터, 웹툰, 패션, 세계관, 디지털디자인, 가격: 44000원
 2. 제목: 인공지능 1 - 현대적 접근방식, 제4판, 가격: 42000원
 3. 제목: 파이썬을 이용한 화학 인공지능, 가격: 39000원
 4. 제목: 국악의 과학적 이해와 인공지능 국악 2 - 정간보 데이터 구축을 위한 정악 모음집, 가격: 35000원
 5. 제목: EU 인공지능법 - EU AI Act, 가격: 34000원
 6. 제목: 인공지능 - 파이썬으로 배우는 머신러닝과 딥러닝, 2판, 가격: 33000원
 7. 제목: ComfyUI 100선 - 설치부터 고급 예제까지, 가격: 33000원
 8. 제목: 인공지능 수학 with 파이썬 - 수학자와 함께하는, 가격: 32000원
 9. 제목: 파이썬으로 만드는 인공지능, 가격: 31000원
 10. 제목: 인공지능 시대의 선형대수학 - 개정3판, 가격: 30000원
- 'ai_top10_books.json' 파일이 생성되었습니다.

```
✓ output  
{} ai_top10_books.json
```

(실시간 데이터이기에 출력 결과가 상이할 수 있음)

생성형 AI를 활용한 키워드 기반 도서 분류하기

목표: 검색한 도서를 생성형 AI를 활용하여 분류하고 파일로 저장합니다.

조건:

1. “fetch_books_by_topic” 함수를 완성하여 '도서'와 관련된 도서 100개를 수집합니다.
2. 검색 결과를 바탕으로 생성형 AI를 활용하여 제목에 '습관', '시간 관리', '독서법', '기타'로 각 도서를 분류합니다.
3. 이 때, 생성형 AI의 응답이 JSON으로 반환되도록 프롬프트 및 파라미터 (temperature or top_p)를 수정합니다.
4. 분류된 데이터를 `output/reading_habits.json` 파일로 저장합니다.

- 결과 화면

```
'reading_habits.json' 파일이 생성되었습니다.
```

```
▼ output  
{ } reading_habits.json
```


TTS를 활용하여 주제별 오디오 파일 생성하기

목표: '음악' 관련 도서 10개를 검색하여 각 도서의 제목, 저자, 그리고 간단한 소개를 포함한 텍스트를 저장하고, 이를 기반으로 오디오 파일을 생성합니다.

조건:

1. 알라딘 API를 활용하여 '음악' 관련 도서를 10개 검색합니다.
2. 각 도서의 제목, 저자, 그리고 책 소개를 추출하여 텍스트 파일(`output/music_books_info.txt`)로 저장합니다.
3. 저장된 텍스트를 읽어 `gTTS` 라이브러리로 오디오 파일(`output/music_books.mp3`)을 생성합니다.

• 결과 화면

```
output/music_books_info.txt 파일이 생성되었습니다.  
output/music_books.mp3 파일이 생성되었습니다.  
모든 작업이 완료되었습니다.
```

```
▼ output  
  ≡ music_books_info.txt  
  🔊 music_books.mp3
```

도전 과제

도전 과제 안내

- 도서 데이터를 생성형 AI를 활용하여 분석 및 요약하고, 요약된 데이터를 오디오로 생성해보기
- 챗봇 동작을 할 수 있도록 전체 코드를 정리해보기

제출

제출 시 주의사항

- 제출기한은 금일 18시까지입니다. 제출기한을 지켜 주시기 바랍니다.
- 반드시 README.md 파일에 단계별로 구현 과정 중 학습한 내용, 어려웠던 부분, 새로 배운 것들 및 느낀 점들을 상세히 기록하여 제출합니다.
 - 단순 완성된 코드만을 나열하지 않습니다.
- 위에 명시된 요구사항은 최소 조건이며, 추가 개발을 자유롭게 진행할 수 있습니다.
- <https://lab.ssafy.com/>에 프로젝트를 생성하고 제출합니다.
 - 프로젝트 이름은 '프로젝트 번호-pjt'로 지정합니다. (ex. 01-pjt)
- 반드시 각 반 담당 강사님을 Maintainer로 설정해야 합니다.