

M6. 데이터 수집 시스템 만들기

3. API를 통한 정기적인 수집 시스템 제작 실습

커리큘럼

과정 목표

- 01 API로 데이터를 수집하는 jupyter notebook 파일을 python 파일로 변환할 수 있다.
- 02 원하는 때에 정기적으로 수집하도록 스케줄러를 구성할 수 있다.
- 03 수집 프로세스가 정상적으로 수행되었는지 확인할 수 있다.

모듈 명	주요 내용
M1. 데이터의 형태	<ul style="list-style-type: none"> 데이터를 정의하고, 강의에서 활용할 데이터에 대한 유형을 소개 JSON, CSV, TSV, XLSX, 등 데이터의 형태를 소개하고 간단한 활용법 학습 Database에 대한 소개 및 SQLite를 활용한 간단한 활용법 숙지
M2. 웹의 기본구조	<ul style="list-style-type: none"> 웹의 개념과 간단한 역사 및 작동원리에 대한 내용을 파악 실제 다양한 웹페이지를 통해 웹을 구성하고 있는 요소(HTML, Javascript, JSON 등)들을 학습
M3. 크롤링을 위한 파이썬 라이브러리	<ul style="list-style-type: none"> 크롤링을 위한 다양한 파이썬 라이브러리에 대한 소개 주피터노트북을 활용하여 각 라이브러리에 대한 실습 진행
M4. 웹페이지를 활용한 데이터 수집	<ul style="list-style-type: none"> 실제 동작하는 웹페이지를 분석하고 수집하여, 원하는 형태의 데이터로 변환하는 실습 진행 대상 : 네이버금융, 네이버뉴스, DART, 증권사이트 등
M5. API를 활용한 데이터 수집	<ul style="list-style-type: none"> 공공기관이나 민간기업에서 제공하고있는 OpenAPI를 활용하여 데이터를 수집하는 실습 진행 대상 : 공공데이터(부동산, 금융), SNS API, 쇼핑몰 API 등
M6. 데이터 수집 시스템 만들기	<ul style="list-style-type: none"> 위에서 작성된 파이썬 코드를 활용하여, 정기적인 수집 시스템을 만드는 실습 진행

*상기 커리큘럼의 내용 및 교육시간은 기관 및 교육과정의 사정에 따라 변동이 가능합니다.

API를 통한 정기적인 수집 시스템 제작 실습



우리가 구성할 수집 시스템

M5 강의에서 실습했던 jupyter notebook 파일 중

[Youtube 금융위원회 채널 동영상의 조회수 및 좋아요 수 저장하기]를

3시간마다 수집하는 시스템 (실습에서는 1분마다 수집하도록 함)

The screenshot shows a Jupyter Notebook titled "m5_2_youtube_api" with a last checkpoint 16 hours ago. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a status bar indicating "Not Trusted" and "Python 3". The notebook contains several code cells:

```
In [4]: # 금융위원회 유튜브 채널
channel_id = 'UCuJz-PXNMdWQr6TNM4htENw'
params = {
    "key": api_key,
    "part": "id,contentDetails",
    "channelId": channel_id,
    "maxResults": 50,
    "fields": "nextPageToken,pageInfo,items(contentDetails/upload)"
}

In [5]: url = 'https://www.googleapis.com/youtube/v3/activities'
res = requests.get(url, headers=headers, params=params)

In [6]: res.json()

In [7]: # 동영상 목록을 리스트로 저장
video_list = []
for video in res.json()['items']:
    video_id = video['contentDetails']['upload']['videoId']
    video_list.append(video_id)

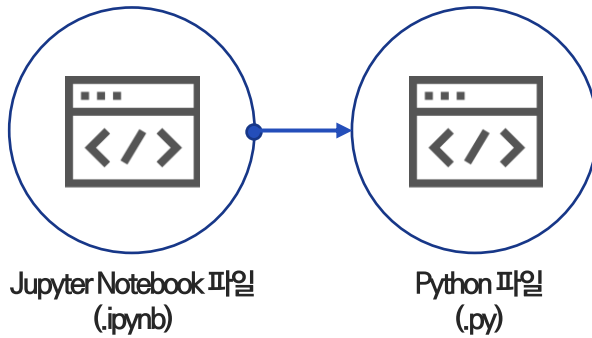
In [15]: params = {
    # api키 항상 먼저 입력
    "key": api_key,
```

API를 통한 정기적인 수집 시스템 제작 실습

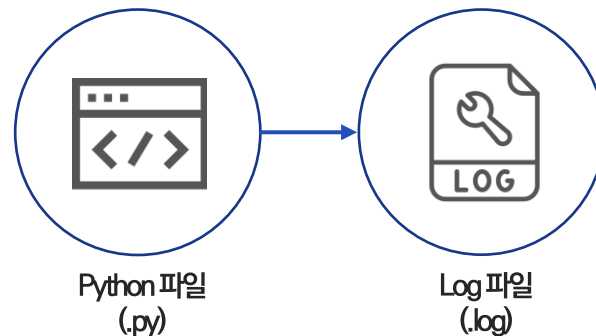


전체 실습 과정 설명

[STEP 1] Jupyter notebook 파일을 하나의 Python 파일로 변경하기



[STEP 2] Python 파일에 실행 결과 로그 파일 기록하는 코드를 추가하기

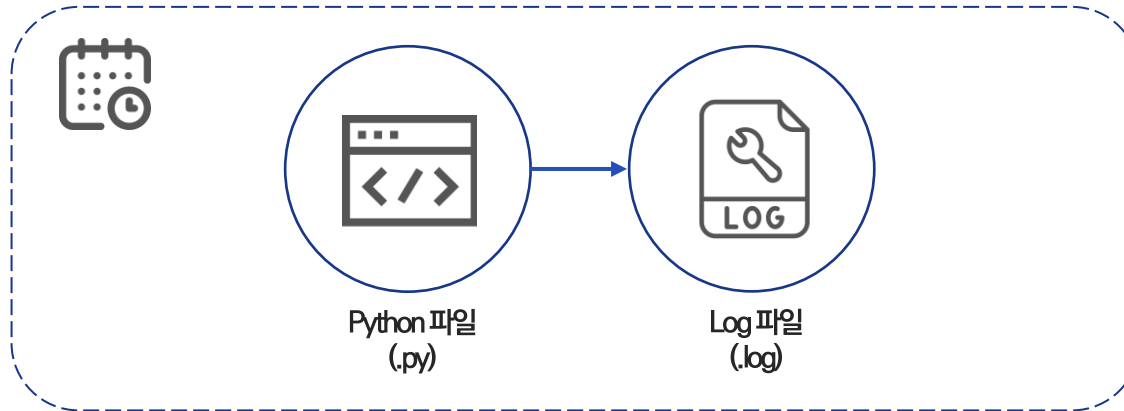


API를 통한 정기적인 수집 시스템 제작 실습

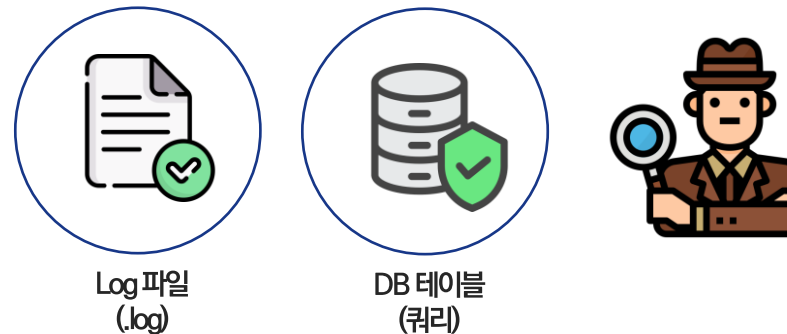


전체 실습 과정 설명

[STEP 3] Python 파일을 스케줄러에 등록하기



[STEP 4] 실행 결과 확인하기



API를 통한 정기적인 수집 시스템 제작 실습



실습을 진행해 봅시다

파이팅!!! 잘 하실겁니다 ☺ 같이 해봐요!