

자기주도 프로젝트

파이선 API 데이터크롤러 구현해보기



계절학기

자기주도 PRJ : Python API Data Crawler 구현

서 성 수 Project consultant

- 삼성 청년 SW 아카데미 프로젝트 컨설턴트
- SA & TA & DA & AA & Development
- 통신3사, 금융권 위주 Dev 20년+
- Oracle Certified Professional
- 쿠크 홈시스 연구원 & Inverter AI 개발
- 컴퓨터공학 전공



Contents

- I 웹크롤러 와 데이타크롤러 의 차이
- II 자기주도학습 과제 개요
- III 과제 목표
- IV 필수지식학습
- V 기능명세



웹크롤러 와 데이터크롤러 의 차이



☑ 개요 설명

- 웹 크롤러 :

불특정 다수의 Web 에 Open된 Data 들을 불특정 다수의 Bot 이 접근하여 수집하는 것

- 데이터 크롤러 :

불특정 다수의 Web 에 승인된 Agent 나 Bot 등의 기능이 접근하여 “API 로 허가된” 데이터들을 수집하는 것

Web Crawler = 데이터 제공자의 의지와 상관없이 “불법의 영역을 침범할수 있는” 수집기능

API Data Crawler = 데이터 제공자로부터 “최소한의 허가를 받은” 수집기능

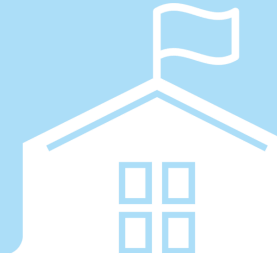
자기주도학습 과제 개요



☑ 개요 설명

- ✔ 본 과제는 Python 의 fastAPI 와 MongoDB를 활용한 간단한 Python API 서버 및 Web API data Crawler 를 만들어보는 실습과제 입니다.
Python 의 API 개발 기초와 NoSQL 사용 기초, Command 사용 기초를 전달하는 것이 주 목적입니다.

과제 목표



☑ 목표 설명

- 1) Python 을 설치한다.
- 2) Python FastAPI 를 세팅한다.
- 3) MongoDB를 설치한다.
- 4) MongoDB 에 조회할 기본 데이터 (상품명) 를 입력한다.
- 5) 네이버 쇼핑 API 에 개발자 등록을 하고 사용키를 취득한다.
- 6) Python FastAPI 에서 MongoDB 의 기본데이터 (상품명) 를 조회한다.
- 7) 조회한 데이터로 네이버 쇼핑 API 를 호출하여 네이버 스토어에서 판매중인 정보를 조회한다.
- 8) 조회한 데이터를 본인이 설치한 MongoDB 에 저장한다.

* 본 과제는 크게, **8단계**의 개발과정을 거치게 됩니다.
많이 들 만들어보는 웹 크롤러와 데이터 크롤러의 차이를 생각하며 개발을 진행해 봅시다.

필수 지식 학습



☑ 목표 설명

- Python fastAPI 로 MongoDB 에 Naver 에서 API 조회한 데이터를 입력하려면 아래 5단계의 진행이 필수입니다.

Req.	Category
1	환경 구축 과제목표 1) python 설치, 과제목표 2) Python FastAPI framework 구성 및 RESTful API 개발, CMD 동작 개발 과제목표 3) MongoDB 설치, 과제목표 5) naver API 요청키 신청
2	과제목표 4) MongoDB 에 default Data insert
3	과제목표 6, 8) Python FastAPI 와 Mongdo DB 연동 및 db select & insert 기능 개발
4	과제목표 7) 서버 실행 및 기능 확인

* 이해를 돕기 위하여 진행 방법 상세하게 업데이트 하였습니다.

기능 명세



☑ Hint

- MongoDB에 입력할 데이터 규격을 소개합니다.
2022년 Bluetooth Headphone 목록을 기준으로 입력데이터의 규격을 설정 해 봤습니다.
해당 규격에 맞도록 MongoDB에 data가 insert 및 save 되는것이 목표입니다.

```
{
  id:ObjectId('635b9503f8c4fca80ad84b86')
  productObj: Object
  prd_id: "WH-1000XM5"
  prd_maker: "Sony"
},
{
  id:ObjectId('635b9503f8c4fca80ad84b87')
  productObj: Object
  prd_id: "Momentum4"
  prd_maker: "Sennhisier"
},
{
  id:ObjectId('635b9503f8c4fca80ad84b89')
  productObj: Object
  prd_id: "QC45"
  prd_maker: "Bose"
}
```


여러분의
폭풍 성장을 응원합니다!!

