

크롤링 함수 실행 방법

크롤링 함수를 Django서버에서 분리시켰습니다. 따라서 크롤링 함수를 따로 실행시켜야합니다.

1st 프로젝트 최상위 폴더에 있는 requirements.txt에서 APScheduler 패키지를 설치합니다.

APScheduler는 파이썬 함수를 주기적으로 실행시켜주는 스케줄러 입니다.

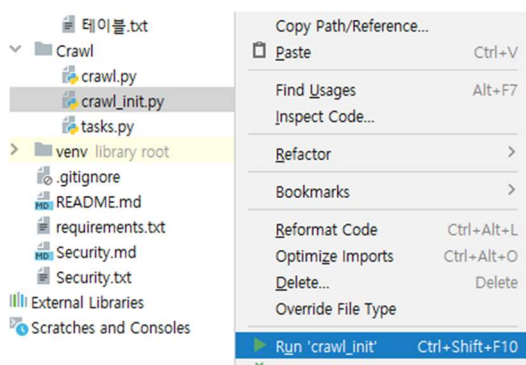
```
konlpy==0.6.0
tqdm==4.65.0
gmail==0.6.3
mysqlclient==2.1.1
PMySQL==1.0.2
APScheduler==3.10.1
```

이때 따로 pip를 하지말고 위에 나타나는 창으로 install을 진행해주세요



Install requirement 클릭 (이때, gensim 설치 오류가 발생하면 그건 제외해서 설치해주세요)

2nd 프로젝트 최상위 폴더/Crawl 폴더에 크롤링 함수들을 배치했습니다.



Crawl_init.py를 실행시키면 전체 게시판 전체 페이지를 크롤링해옵니다.

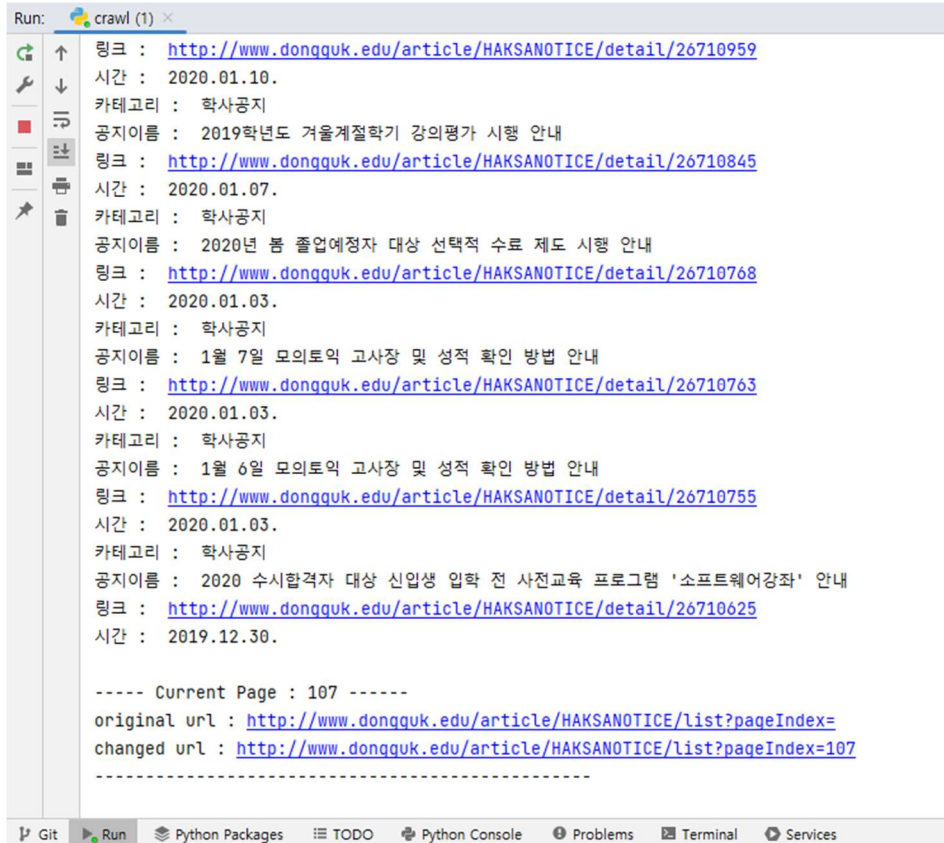
pull하시고 한번만 실행해주세요. 전체 크롤링하는데 3시간 걸립니다.

실행되는동안 프로젝트 진행해주세요

Tasks.py를 실행시키면 크롤링과 주기 업데이트를 주기적으로 실행시킵니다

3rd 작업 내용 확인하기

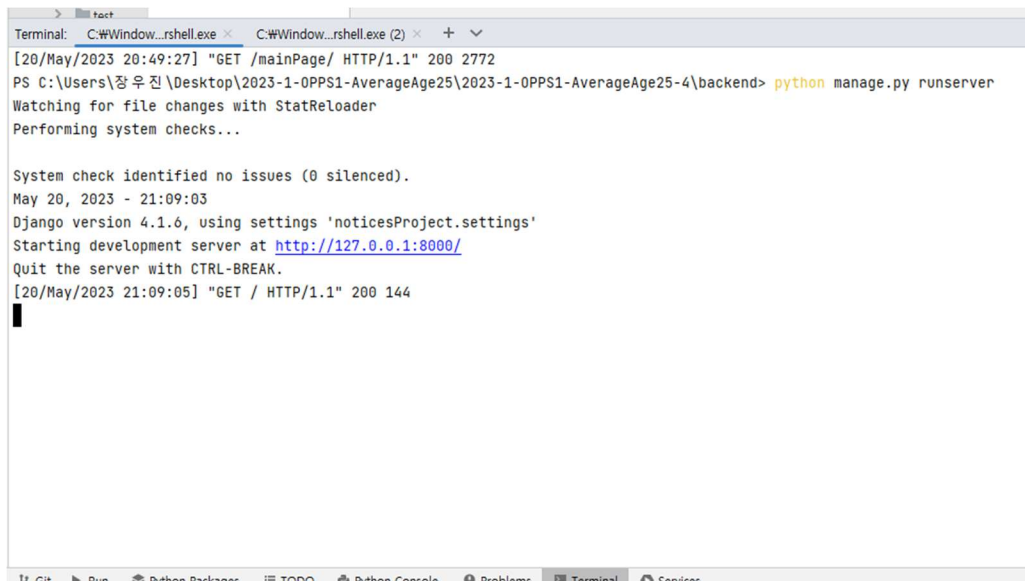
크롤링 코드를 Django 서버에서 분리하였기에, 이제는 crawl코드 동작을 Run창에서 확인할 수 있습니다.



```
Run: crawl (1) ×
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710959
시간 : 2020.01.10.
카테고리 : 학사공지
공지이름 : 2019학년도 겨울계절학기 강의평가 시행 안내
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710845
시간 : 2020.01.07.
카테고리 : 학사공지
공지이름 : 2020년 봄 졸업예정자 대상 선택적 수료 제도 시행 안내
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710768
시간 : 2020.01.03.
카테고리 : 학사공지
공지이름 : 1월 7일 모의토익 고사장 및 성적 확인 방법 안내
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710763
시간 : 2020.01.03.
카테고리 : 학사공지
공지이름 : 1월 6일 모의토익 고사장 및 성적 확인 방법 안내
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710755
시간 : 2020.01.03.
카테고리 : 학사공지
공지이름 : 2020 수시합격자 대상 신입생 입학 전 사전교육 프로그램 '소프트웨어강좌' 안내
링크 : http://www.dongguk.edu/article/HAKSANOTICE/detail/26710625
시간 : 2019.12.30.

----- Current Page : 107 -----
original url : http://www.dongguk.edu/article/HAKSANOTICE/list?pageIndex=
changed url : http://www.dongguk.edu/article/HAKSANOTICE/list?pageIndex=107
-----
```

기존 Django 서버 동작을 확인하려면 Terminal에서 확인할 수 있습니다.



```
Terminal: C:\Window...rshell.exe × C:\Window...rshell.exe (2) × + v
[20/May/2023 20:49:27] "GET /mainPage/ HTTP/1.1" 200 2772
PS C:\Users\장우진\Desktop\2023-1-OPPS1-AverageAge25\2023-1-OPPS1-AverageAge25-4\backend> python manage.py runserver
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).
May 20, 2023 - 21:09:03
Django version 4.1.6, using settings 'noticesProject.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
[20/May/2023 21:09:05] "GET / HTTP/1.1" 200 144
```