



Tut 1: Hướng dẫn sử dụng Scraper

- Tác giả: Mai Đức Trung.
- Nhóm nghiên cứu về AI của Trường Đại học Bách Khoa Thành phố Hồ Chí Minh.

1. Cài đặt môi trường chạy:

a. MongoDB

<https://www.mongodb.com/>

b. Python 3

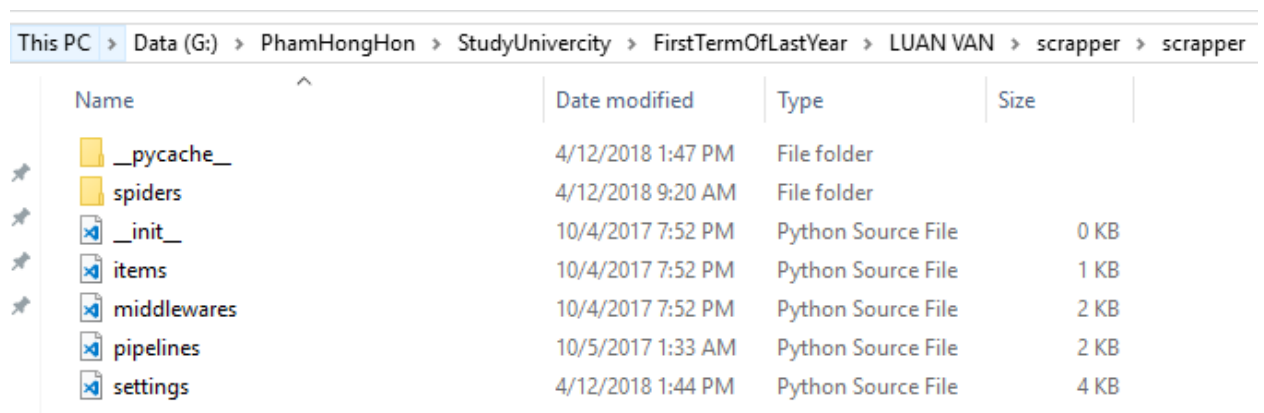
<https://www.python.org/>

c. Thư viện Scrapy trong Python 3

<https://doc.scrapy.org/en/latest/intro/install.html>

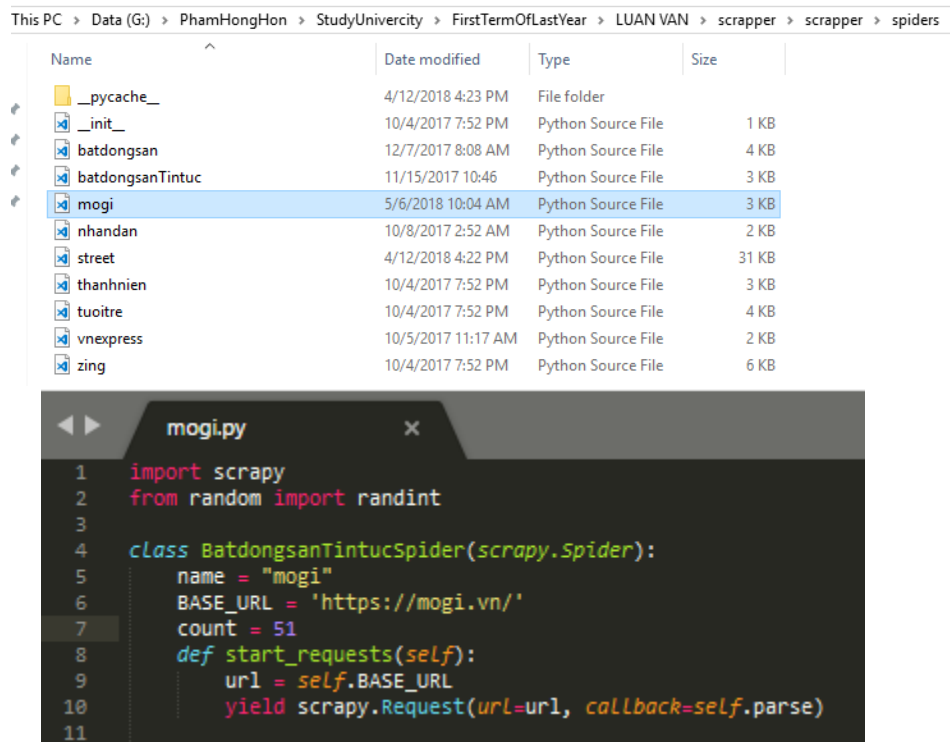
2. Hướng dẫn chạy:

- B1: Chạy MongoDB (ở dưới là hướng dẫn)
<https://o7planning.org/vi/10265/huong-dan-cai-dat-va-cau-hinh-database-mongodb>
- B2: Mở thư mục như trong hình dưới:
Và mở CMD python tại thư mục này.



This PC >	Data (G:) >	PhamHongHon >	StudyUnivercity >	FirstTermOfLastYear >	LUAN VAN >	scrapper >	scrapper
Name	Date modified	Type	Size				
__pycache__	4/12/2018 1:47 PM	File folder					
spiders	4/12/2018 9:20 AM	File folder					
__init__.py	10/4/2017 7:52 PM	Python Source File	0 KB				
items.py	10/4/2017 7:52 PM	Python Source File	1 KB				
middlewares.py	10/4/2017 7:52 PM	Python Source File	2 KB				
pipelines.py	10/5/2017 1:33 AM	Python Source File	2 KB				
settings.py	4/12/2018 1:44 PM	Python Source File	4 KB				

- B3: chạy lệnh : scrapy crawl mogi
Trong đó mogi là tên được đặt trong file mogi.py ở trong thư mục spiders.



- Dữ liệu sẽ được lưu trong MongoDB sau khi chạy.
- Chỉ cần sửa những file như mogi.py thành cái của mình, đổi tên là được

3. Các tài liệu tham khảo:

<https://doc.scrapy.org/en/latest/intro/tutorial.html>