**LAB 5**

# Thông tin chung

* Thực hành chương 3: Crawling website, hoàn thiện search engine với dữ liệu bài báo đã crawl được từ các webstie.
* Sinh viên lưu file với tên theo định dạng sau: Lab5\_MSSV.ipynb,
* Sinh viên nộp lên Elearning của lớp học tại buổi thực hành 5.
* Deadline: 17h00, ngày 14/03/2023

# Yêu cầu

* Sử dụng các IDE (anaconda, pycharm, google colab,…) để tạo file thực hành.

# Nội dung

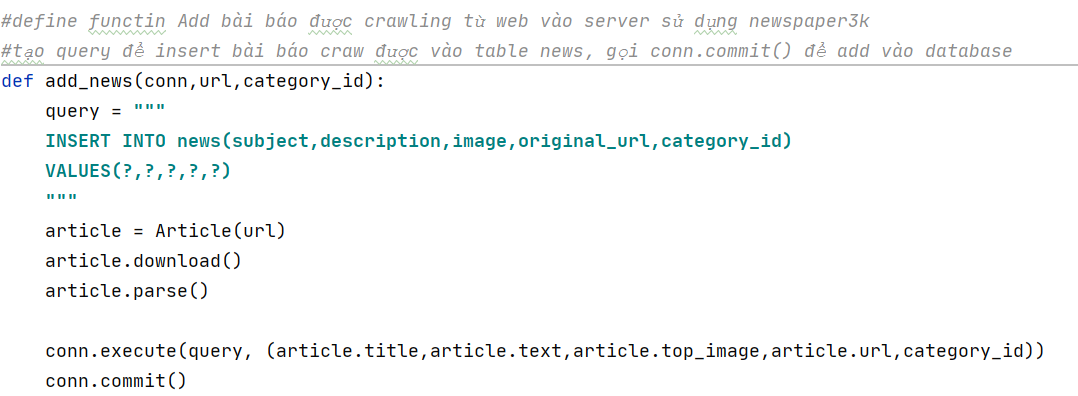
### Chuẩn bị

* 1. **Thực hiện**

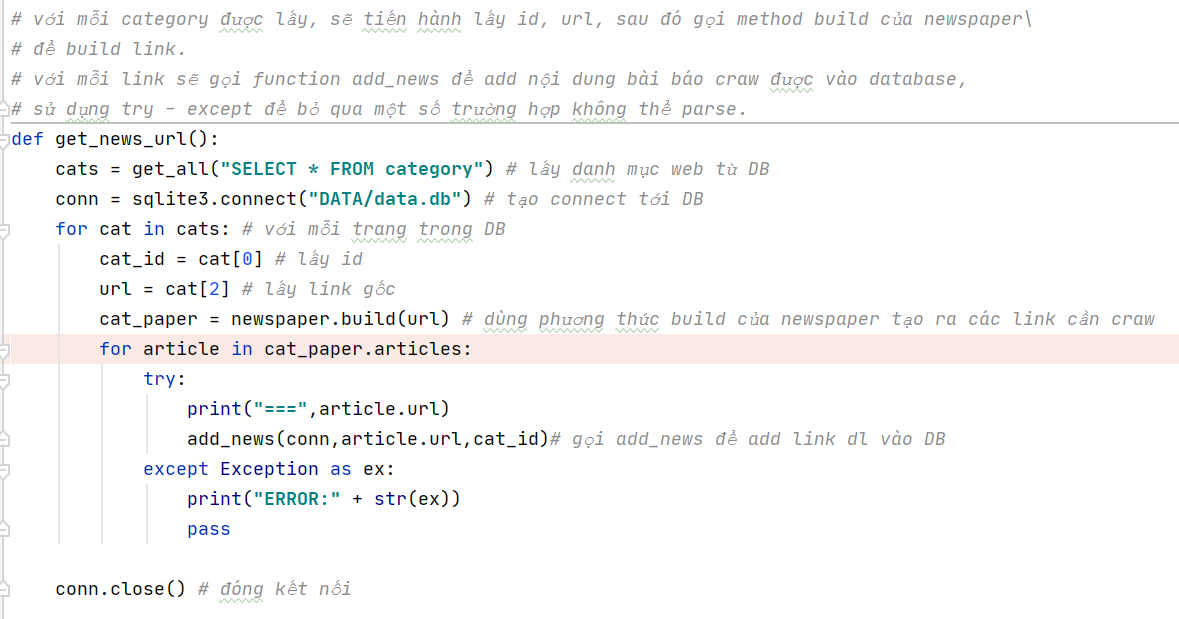
B1. Tạo new directory trong project lab 4, đặt tên directory là json\_file.

***B2. Module utils.py, tạo các function như sau:***

* **add\_news(): để thêm dữ liệu bài báo thu thập được vào database**



* **get\_news\_url(): để build link chi tiết các bài báo từ trang chủ các website được lưu trong category**

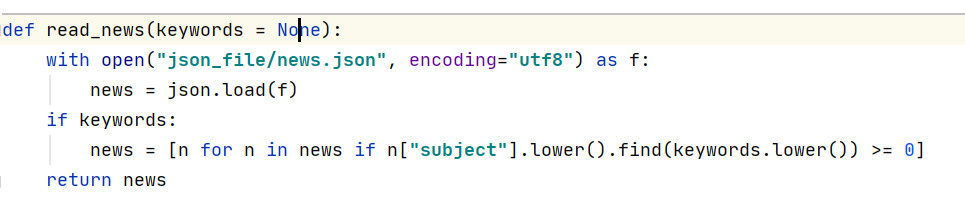
****

***B3. Module render\_templates.py, tạo các function như sau:***

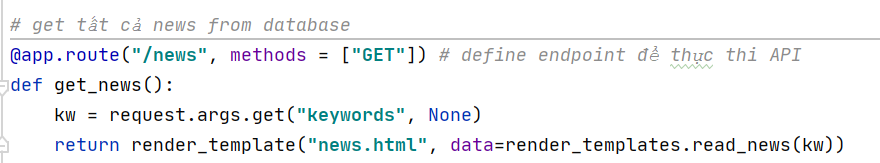
* **get\_news(): load các news từ database ghi vào file json: news.json trong directory json\_file.**



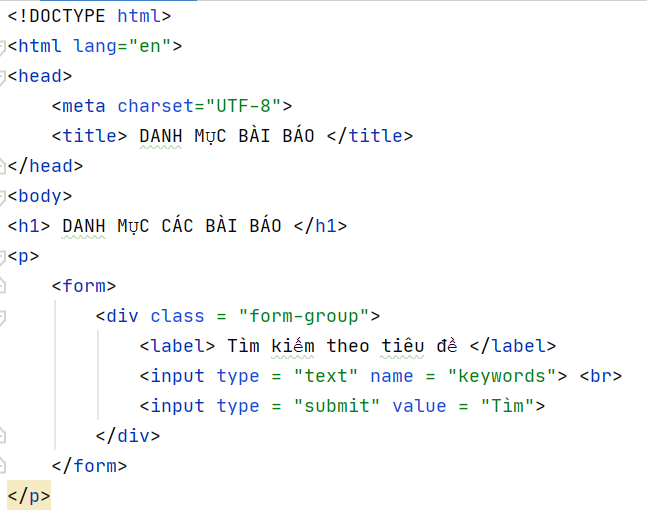
* **read\_news(keyword = none): đọc các news từ file json news.json để tạo ra dữ liệu cho news.html**

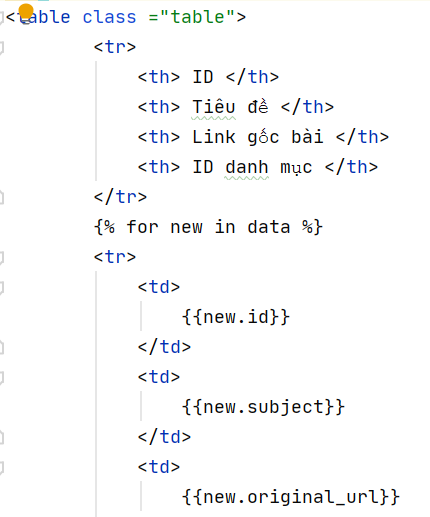


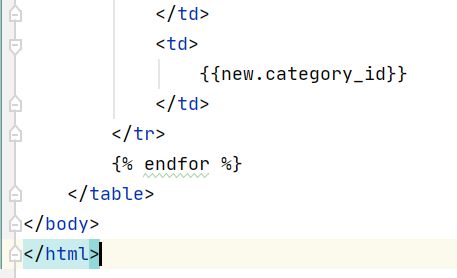
***B4. Module api.py: define endpoint để render dữ liệu từ news.html thành dữ liệu website, nếu không có keyword sẽ load tất cả news, nếu có keyword sẽ load các news chứa keyword tương ứng***

****

**B5: Thiết lập file news.html**

****

****

****