|  |  |  |  |
| --- | --- | --- | --- |
| **THÔNG TIN THÀNH VIÊN NHÓM** | | | |
| **#** | **Họ tên** | **MSSV** | **Lớp** |
| 1. | Lê Khôi Đệ | 1411010102 | 15DTH06 |
| 2. | Nguyễn Bá Tâm | 1611061022 | 16DTHC2 |
| 3. |  |  |  |

**NỘI DUNG BÁO CÁO TUẦN 4**

Báo cáo các nội dung bao gồm:

* Nội dung đã thực hiện: Thực hành được Crawl Data từ toidicodedao
* Nội dung đang thực hiện: Phân tích và nghiên cứu các bài tham khảo để có thêm nhiều thông tin về kỹ thuật Crawl Data
* Khó khăn: Hiện tại cá nhân đang phục hồi sức khỏe sau khi bị sốt siêu vị nên báo cáo tương đối ngắn mong cô thông cảm.
* Câu hỏi: Em không có câu hỏi gì cả

**4.1 Dữ liệu khi cào về được Import vào Database(Cập nhật liên tục)**

<https://toidicodedao.com/2015/07/28/tutorial-trich-xuat-thong-tin-tu-website-voi-html-aglitity-pack/>

LÀM MỘT DỰ ÁN MẪU TỪ toidicodedao

Mục đích: thực hành được kỹ thuật Crawl Data dùng ngôn ngữ C#

Ngôn ngữ sử dụng: HTML AgilityPack, Fizzler, C#

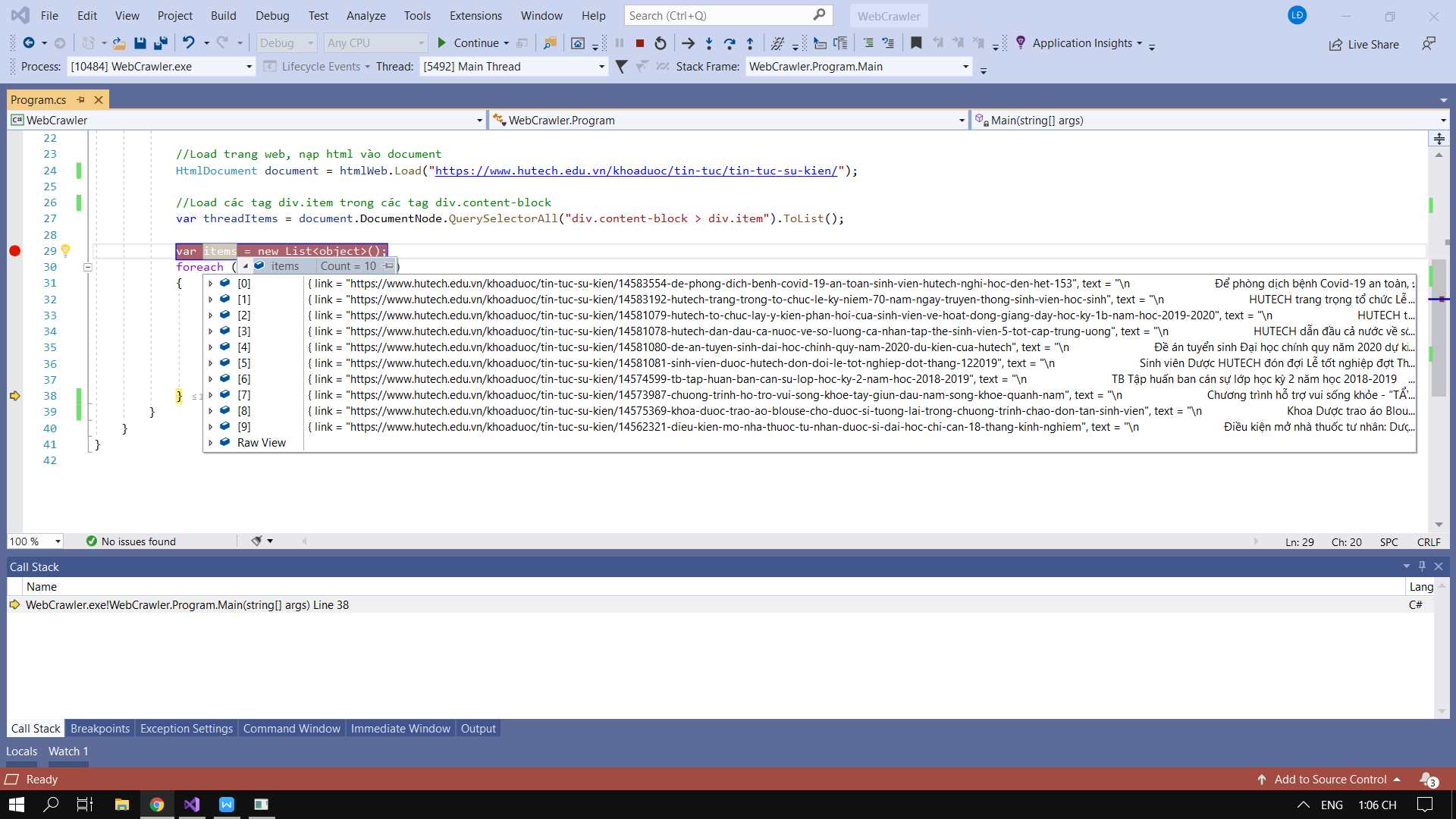
Phần mềm sử dụng: Visual Studio 2019

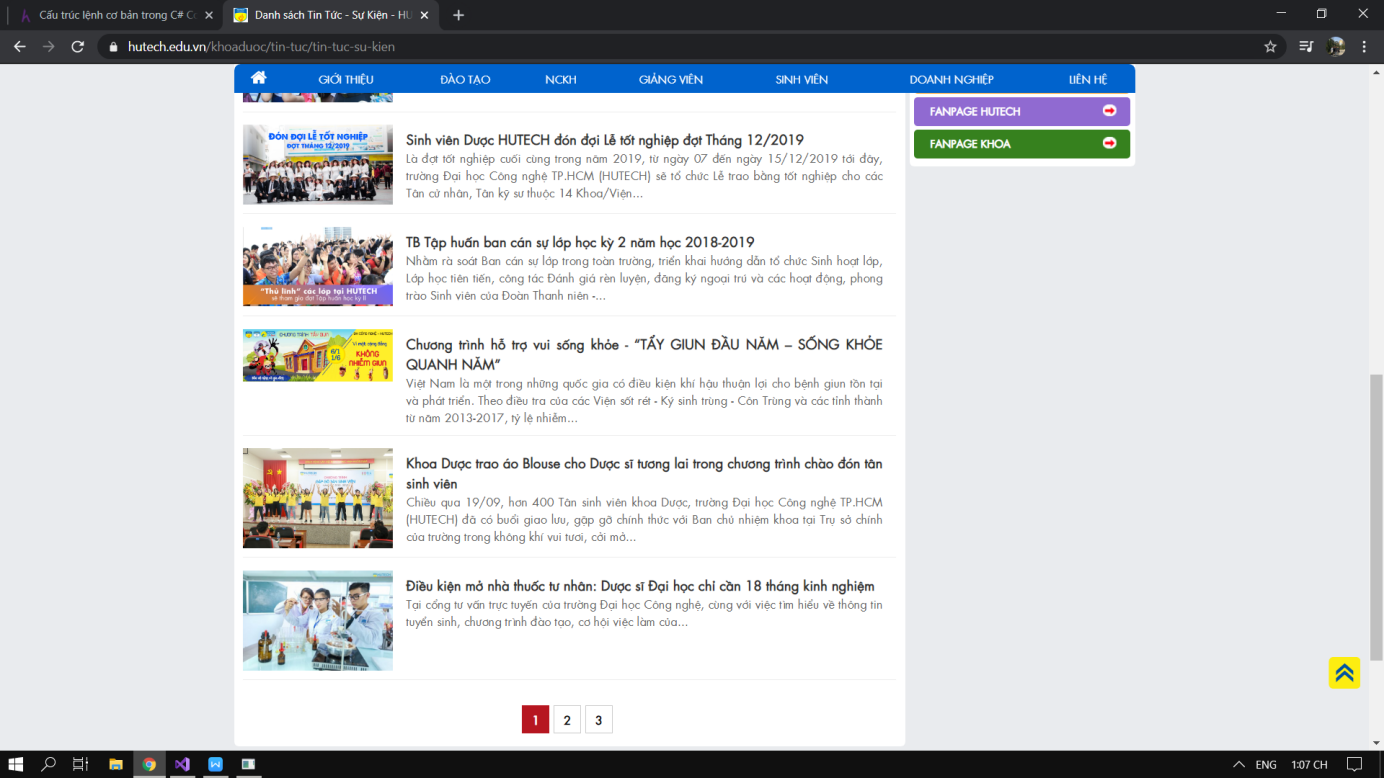
Làm được:

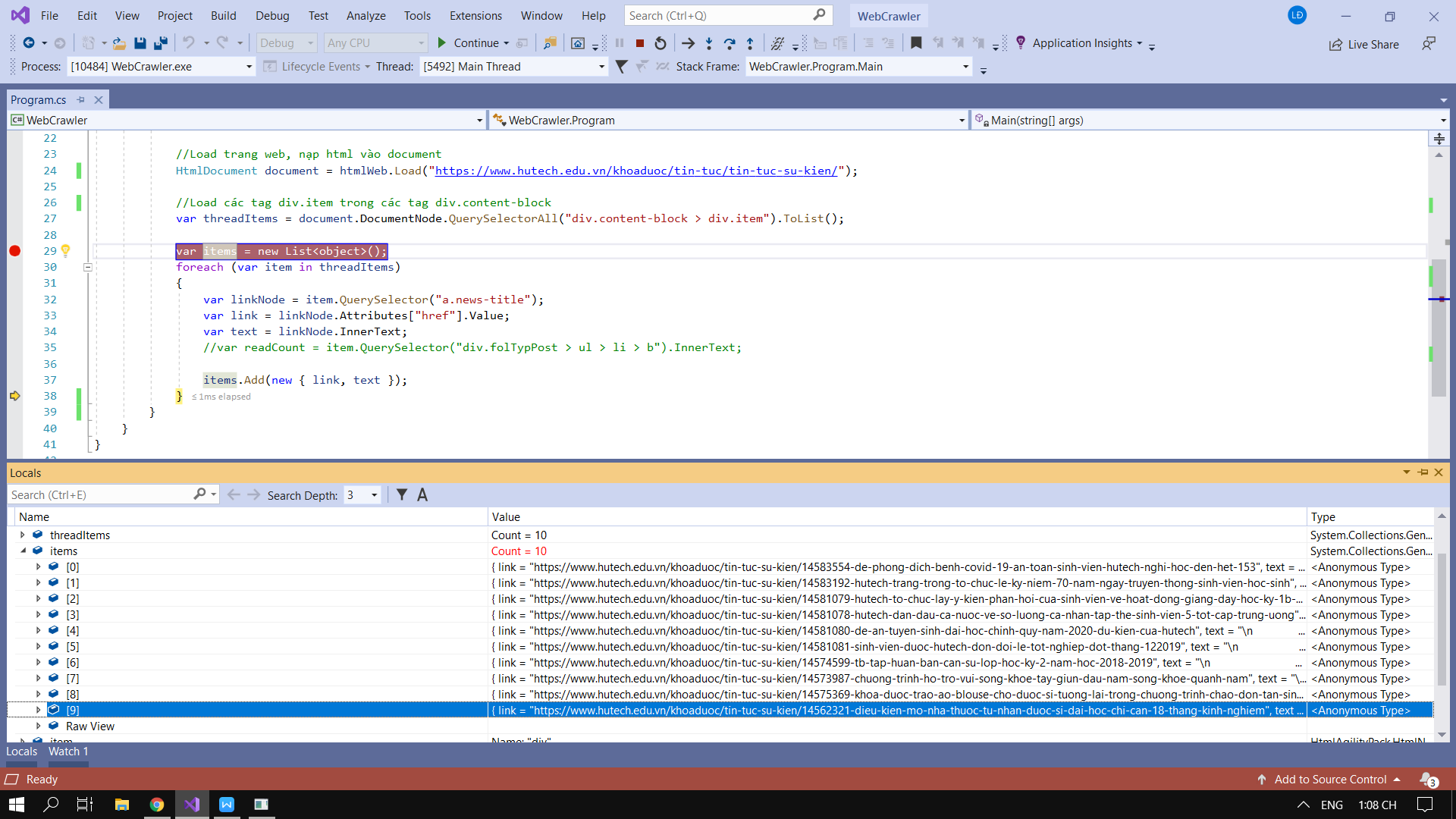
Crawl được dữ liệu (<https://www.hutech.edu.vn/khoaduoc/tin-tuc/tin-tuc-su-kien>)

Phân tích được điểm sai khi phân tích cấu trúc HTML hồi tuần 3

Chưa làm được: Suy nghĩ phương pháp thêm nhiều đường dẫn khác để lấy dữ liệu, suy nghĩ ra phương pháp tạo objets để import vào Database, lấy được dữ liệu khác khi qua phân trang mới







Thấy ở đây items [9] “Điều kiện mở nhà thuốc…” khớp với tiêu đề sự kiện ở trên trang Website Hutech đường dẫn đang lấy.