|  |  |  |  |
| --- | --- | --- | --- |
| **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 3**

Báo cáo các nội dung bao gồm:

* Nội dung đã thực hiện: Thực hành sử dụng kỹ thuật Crawl Data từ HowKteam
* 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: Dự án mẫu Fail chi tiết nằm ở mục 3.1
* Câu hỏi: Em không có câu hỏi gì cả
* Trả lời câu hỏi tuần 2: Việc “trích xuất thông tin” từ cô Thơ đề xuất nhóm chúng em đã họp và đưa ra quyết định là chỉ tìm hiểu cho biết và không thực hiện.Hiện tại nhóm chỉ muốn tập trung toàn bộ thời gian và sức lực cho kỹ thuật Crawl Data.

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

<https://www.youtube.com/watch?v=l8zPH8r9nKU&list=PL33lvabfss1wqC6WLIEI46riLZMdvv9UV>

LÀM MỘT DỰ ÁN MẪU TỪ HowKteam

Mục đích: thực hành được kỹ thuật Crawl Data từ hướng đi của HowKteam

Ngôn ngữ sử dụng: C#, RegEx

Phần mềm sử dụng: Visual Studio 2019, RegEx Tester 3.2.0

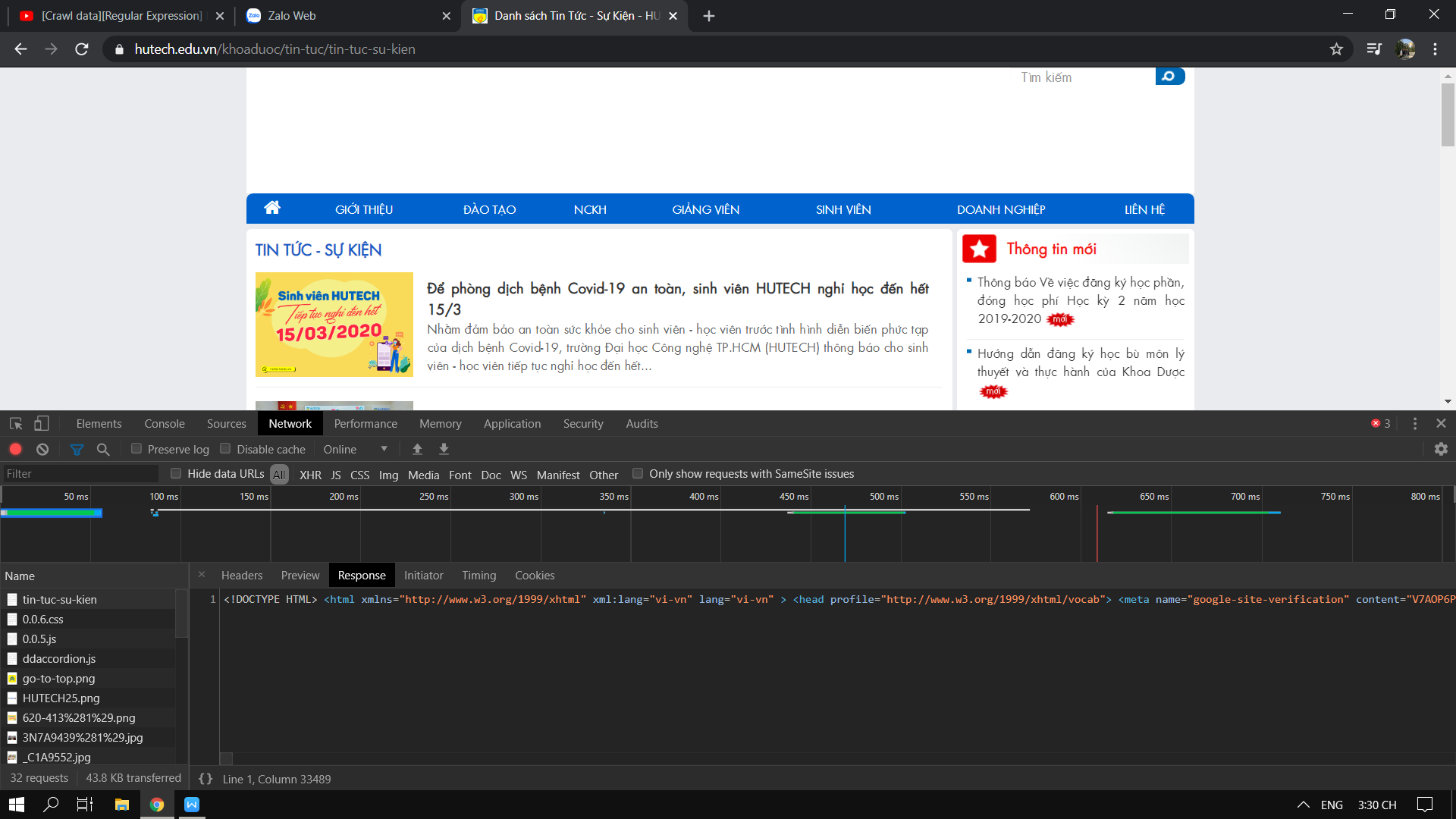
Làm được: Phân tích được cấu trúc HTML từ đường dẫn Website Hutech (<https://www.hutech.edu.vn/khoaduoc/tin-tuc/tin-tuc-su-kien>) trên RegEx Tester. Crawl được cấu trúc HTML

Chưa làm được: Chưa làm được bài 4 của khóa học từ HowKteam là bóc tách được dữ liệu, dữ liệu khi trả về luôn ra Null

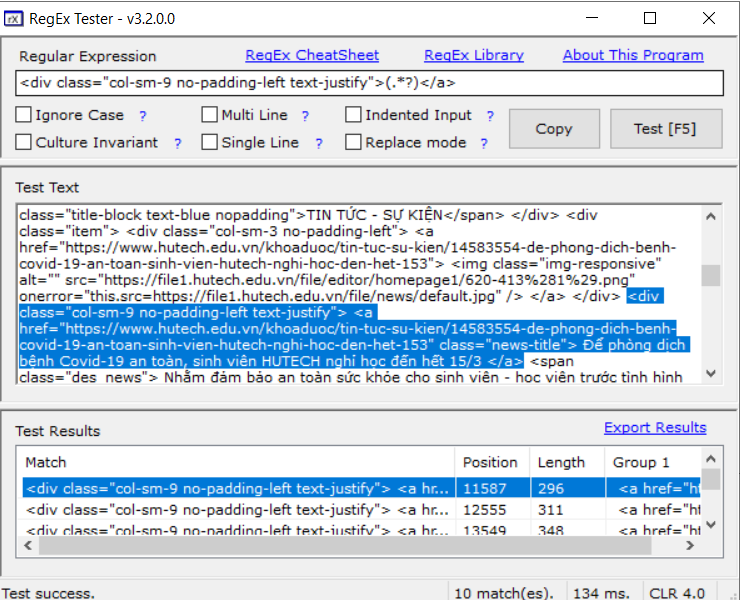
Khó khăn: Khi phân tích xong được Website cần lấy, thì việc chọn ngôn ngữ sử dụng và phần mềm gì để viết là một vấn đề hơi khó khăn làm sao để hiểu rõ về nó để áp dụng vào hệ thống. Tóm lại từ hướng đi chọn ngôn ngữ sử dụng và phần mềm để viết đang cần thời gian để phân tích thêm bằng cách đọc thêm những bài tham khảo và thực hành dự án mẫu

Những kết quả gặt hái được:

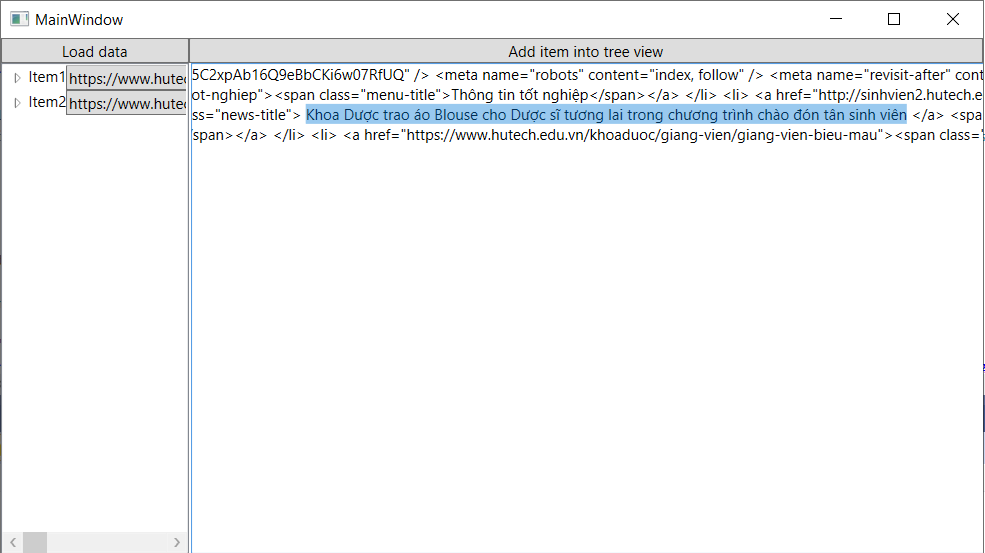
Lấy toàn bộ cấu trúc HTML đưa vào RegEx Tester



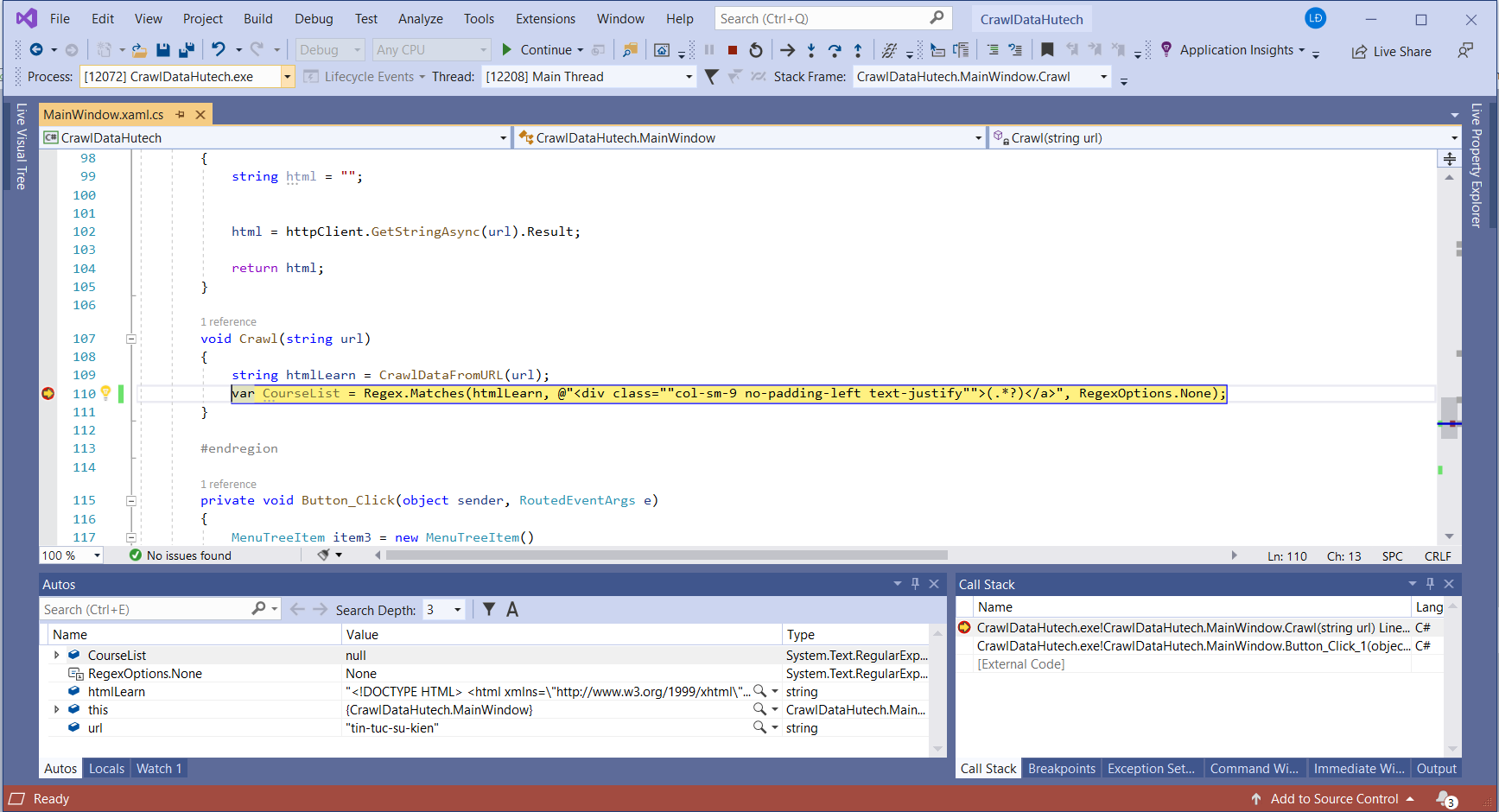
Bóc tách được dữ liệu cần lấy trên RegEx Tester



Crawl được cấu trúc HTML của đường dẫn cần bóc tách



Khi chạy code việc bóc tách dữ liệu. Dữ liệu ở đây là tên sự kiện thì kết quả trả về Null nếu ra đúng thì ra là 10



CourseList là null.

Lỗi có thể nằm ở việc dòng code RegEx không đúng

Hay lỗi có thể nằm ở việc phân tích sai trên RegEx Tester

Quan trọng vấn đề không báo lỗi mà vẫn chạy.