

ĐỒ ÁN TỔNG HỢP

XÂY DỰNG CÔNG THÔNG TIN SỰ KIỆN TẠI ĐẠI HỌC CÔNG NGHỆ - HUTECH

Ngành: **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **CÔNG NGHỆ PHẦN MỀM**

Giảng viên hướng dẫn : TS. Lê Thị Ngọc Thơ

Sinh viên thực hiện : **Lê Khôi Đệ**

MSSV: 1411010102 Lớp: 15DTH06

Sinh viên thực hiện : **Nguyễn Bá Tâm**

MSSV: 1611061022 Lớp: 16DTHC2

TP. Hồ Chí Minh, 2020

LỜI CAM ĐOAN

Nhóm em xin cam đoan những điều được viết trong báo cáo đều là những thành quả, quá trình lao động không ngừng nghỉ trong 12 tuần. Điều là những ý tưởng sơ lược được nhóm em nghĩ ra và nghiên cứu.

Đề tài “Xây dựng cổng thông tin các thông tin tại Đại học Công Nghệ TP.HCM - HUTECH” một ý tưởng xuất phát điểm là từ bản thân. Vì trong quá trình học tập và rèn luyện tại trường thì Điểm rèn luyện luôn chiếm một vai trò trong việc đánh giá một sinh viên. Sau khảo sát 100 bạn tại trường Đại Học Công Nghệ với câu hỏi “Bạn thường tham gia các hội thảo, sự kiện tại trường qua đâu?” thì 70% các bạn đều trả lời là qua mạng xã hội, 20% là qua các băng rôn được treo tại trường, còn lại là truyền miệng.

Tại sao đa số các sự kiện của trường đều được biết qua mạng xã hội, còn trang Website của trường lại không được cập nhật đầy đủ để sinh viên có thể theo dõi. Để trả lời câu hỏi trên thì nhóm em đã đi tìm hiểu và nghiên cứu thì thấy được rất nhiều vấn đề.

Trang Website của trường thường chỉ cập nhật các sự kiện tổ chức hàng năm hay các sự kiện nổi tiếng như Hutech Got Talent,. Và mỗi sự kiện từng khoa thường nằm trong từng danh mục cho thấy thiếu thân thiện và thiếu tiện lợi. Còn thông qua mạng xã hội thì từng trang riêng lẻ, nhóm riêng lẻ hay bị bỏ sót nếu không chịu cập nhật thường xuyên.

Xuất phát từ những lý do trên, cùng với sự hướng dẫn của cô Lê Thị Ngọc Thơ, nhóm em thực hiện đề tài “Xây dựng cổng thông tin các thông tin tại đại học công nghệ (Hutech)” để vận dụng những kiến thức đã được học vào thực tiễn, xây dựng một hệ thống cập nhật sự kiện các thông tin tại trường Hutech.

MỤC LỤC

LỜI CAM ĐOAN.....	2
MỤC LỤC.....	3
DANH MỤC CÁC HÌNH VẼ.....	5
CHƯƠNG 1 TỔNG QUAN.....	7
1.1 Giới thiệu đề tài.....	7
1.2 Cấu trúc phân rã công việc (WBS).....	9
1.2.1 Xác định (10/02/2020 - 29/02/2020).....	9
1.2.2 Phân tích và thiết kế (1/03/2020 - 04/04/2020).....	9
1.2.3 Thực hiện (05/04/2020 - 23/05/2020).....	10
1.2.2 Kiểm thử (24/05/2020 - 30/05/2020).....	10
1.2.3 Kết luận (31/05/2020 - 27/06/2020).....	11
1.3 Cấu trúc đồ án.....	12
CHƯƠNG 2 CƠ SỞ LÝ THUYẾT.....	13
2.1 Thực trạng vấn đề.....	13
2.2 Giới thiệu công nghệ.....	14
2.2.1 Công nghệ ASP.NET.MVC 5 của Microsoft là gì?.....	15
2.2.2 Công nghệ SQL Server của Microsoft.....	16
2.2.3 Khái niệm Crawl Data.....	17
CHƯƠNG 3 : PHÂN TÍCH VÀ THIẾT KẾ.....	18
3.1 Đặc tả yêu cầu.....	18
3.1.1 Quản lý sự kiện.....	19
3.1.2 Quản lý bộ phận.....	20
3.1.3 Quản lý cơ sở.....	21
3.1.4 Admin đăng nhập.....	21
3.1.5 Danh mục quận huyện.....	22
3.2 Mô hình.....	23
3.2.1 Lược đồ UseCase.....	23
3.2.2 Class Diagram.....	26
3.2.3 ERD.....	26
3.3 Crawl Data.....	27
3.3.1 Những thông tin cần trích xuất trên trang web.....	27

3.3.2 Phân tích cấu trúc trang web.....	27
3.3 Giao diện.....	36
3.3.1 Giao diện trang chủ.....	36
3.3.2 Giao diện xem chi tiết sự kiện.....	37
CHƯƠNG 4: CÀI ĐẶT VÀ ĐÁNH GIÁ KẾT QUẢ.....	38
4.1 Crawl Data.....	38
4.1.1 Cài đặt thư viện.....	38
4.1.2 Chương trình Crawl Data.....	39
4.1.3 Đánh giá kết quả.....	40
4.2 Nạp JSON vào SQL đầy lên hệ thống.....	41
4.3 Chức năng của hệ thống.....	44
4.3.1 Đăng nhập Admin.....	44
4.3.2 CRUD Quản lý sự kiện.....	47
4.3.3 CRUD Quản lý bộ phận.....	50
4.3.4 CRUD Quản lý cơ sở.....	52
4.3.5 CRUD Quản lý danh mục quận huyện.....	54
CHƯƠNG 5 KẾT LUẬN.....	56
4.1 Kết luận chung.....	56
4.2 Khẳng định kết quả đạt được.....	56
4.2.1 Kết quả đạt được.....	56
4.2.2 Đóng góp.....	57
4.3. Đánh giá phần mềm.....	59
4.2.1 Ưu điểm.....	59
4.2.2 Nhược điểm.....	59
4.4. Hướng phát triển và mở rộng đề tài.....	59
TÀI LIỆU THAM KHẢO.....	60

DANH MỤC CÁC HÌNH VẼ

Hình 3.1	UseCase tổng thể.....	23
Hình 3.2	UseCase quản lý bộ phận.....	23
Hình 3 .3	UseCase quản lý cơ sở.....	24
Hình 3.4	UseCase quản lý danh mục quận huyện.....	24
Hình 3.5	UseCase quản lý sự kiện.....	25
Hình 3.6	Class Diagram tổng thể.....	26
Hình 3.7	ERD tổng thể.....	26
Hình 3.8	Phân tích Website định cào.....	27
Hình 3.9	Phân tích Website định cào 1.....	28
Hình 3.10	Phân tích Website định cào 2.....	28
Hình 3.11	Phân tích Website định cào 2.....	29
Hình 3.12	Phân tích Website định cào 3.....	29
Hình 3.13	Phân tích Website định cào 4.....	30
Hình 3.14	Phân tích Website định cào 5.....	31
Hình 3.15	Phân tích Website định cào 6.....	32
Hình 3.16	Phân tích Website định cào 7.....	32
Hình 3.17	Phân tích Website định cào 8.....	33
Hình 3.18	Phân tích Website định cào 9.....	34
Hình 3.19	Phân tích Website định cào 10.....	34
Hình 3.20	Phân tích Website định cào 11.....	35
Hình 3.21	Phân tích Website định cào 12.....	35
Hình 3.22	Giao diện trang chủ người dùng.....	36
Hình 3.23	Giao diện trang chủ người dùng 1.....	36
Hình 3.24	Giao diện trang chủ người dùng 2.....	37
Hình 3.25	Giao diện xem chi tiết sự kiện người dùng.....	37
Hình 4.26	Cài đặt môi trường Crawl Data.....	38
Hình 4.27	Cài đặt môi trường Crawl Data 1.....	38
Hình 4.28	Kết quả Crawl Data.....	40
Hình 4.29	Câu lệnh SQL đưa dữ liệu từ JSON vào hệ thống.....	41
Hình 4.30	Dữ liệu từ JSON đã lên hệ thống.....	41
Hình 4.31	Câu lệnh SQL đưa dữ liệu từ JSON lên hệ thống 1.....	42

Hình 4.32	Câu lệnh SQL đưa dữ liệu từ JSON lên hệ thống 2.....	42
Hình 4.33	Dữ liệu từ JSON đã lên hệ thống 1.....	43
Hình 4.34	Giao diện đăng nhập Admin 01.....	44
Hình 4.35	Giao diện đăng nhập Admin 02.....	44
Hình 4.36	Giao diện đăng nhập Admin 03.....	45
Hình 4.37	Giao diện đăng nhập Admin 04.....	45
Hình 4.38	Giao diện khi đăng nhập tài khoản thành công.....	46
Hình 4.39	Giao diện danh sách sự kiện admin.....	47
Hình 4.40	Giao diện tạo sự kiện.....	47
Hình 4.41	Giao diện xem chi tiết sự kiện admin.....	48
Hình 4.42	Giao diện sửa sự kiện admin.....	48
Hình 4.43	Giao diện xóa sự kiện admin.....	49
Hình 4.44	Giao diện danh sách bộ phận admin.....	50
Hình 4.45	Giao diện chi tiết bộ phận admin.....	50
Hình 4.46	Giao diện sửa bộ phận admin.....	51
Hình 4.47	Giao diện xóa bộ phận admin.....	51
Hình 4.48	Giao diện danh sách cơ sở admin.....	52
Hình 4.49	Giao diện xem chi tiết cơ sở admin.....	52
Hình 4.50	Giao diện sửa cơ sở admin.....	53
Hình 4.51	Giao diện xóa cơ sở admin.....	53
Hình 4.52	Giao diện danh mục quận huyện admin.....	54
Hình 4.53	Giao diện chi tiết quận huyện admin.....	54
Hình 4.54	Giao diện sửa quận huyện admin.....	55
Hình 4.55	Giao diện xóa quận huyện admin.....	55

CHƯƠNG 1 TỔNG QUAN

1. 1 Giới thiệu đề tài

Ở mỗi trường đại học, đều có một trang Website để giới thiệu, hỗ trợ và cập nhật các thông tin về trường. Là nơi các bạn sinh viên cập nhật thông tin về vấn đề học tập, các sự kiện được tổ chức. Nhưng vấn đề ở đây về các thông tin sự kiện nói riêng ở trường Đại học Công Nghệ Hutech lại được cập nhật quá nhiều kênh khác nhau. Kênh ở đây là các trang mạng xã hội, trang Website trường, các băng rôn, các giảng viên và các bạn sinh viên đại diện ở mỗi lớp.

Theo khảo sát 100 bạn sinh viên tại trường Đại học Công Nghệ Hutech qua 2 câu hỏi:

- Câu hỏi thứ nhất “Bạn thường tham gia sự kiện với mục đích gì?”: thì đa số là muốn tích lũy điểm rèn luyện, tìm hiểu và bám sát các hướng đi của doanh nghiệp, giao lưu học hỏi và giải trí để giảm áp lực học tập.
- Câu hỏi thứ hai “Bạn thường cập nhật thông tin sự kiện trường qua đâu?”: thì thường thông qua mạng xã hội nhưng đôi khi hay bị bỏ sót, vì ít khi cập nhật hoặc bị loãng ở trang bảng tin, chỉ biết các sự kiện của khoa mình thông qua giảng viên môn học và bạn sinh viên đại diện mỗi lớp.

Nhận thấy hầu hết các thông tin sự kiện : 11 phòng- ban, 12 khoa, 8 viện, 7 trung tâm, nhiều câu lạc bộ và nhiều nhóm khác. Điều nằm riêng lẻ ở các trang mạng xã hội, các nhóm tập thể và cũng chỉ số ít có liên quan trong các nhóm hay các trang mạng xã hội đó mới cập nhật được thông tin sự kiện còn phần đông chỉ biết qua các giảng viên môn học và các bạn sinh viên đại diện mỗi lớp.

Với vấn đề cập nhật quá nhiều kênh như vậy, thì nhóm chúng em quyết định xây dựng một trang Website tổng hợp các thông tin sự kiện của trường nói chung và các thông tin sự kiện câu lạc bộ, nhóm nhỏ nói riêng với tiêu chí :

- Đầy đủ thông tin đến các bạn sinh viên.
- Cập nhật nhanh chóng các sự kiện mới nhất.
- Các sự kiện phải chính xác và xác thực.

Project Charter

Project Title : Xây dựng cổng thông tin sự kiện tại Đại học Công Nghệ - Hutech

Project Start Date : 10/02/2020 Projected Finish Date : 09/05/2020

Project Manager : Lê Khôi Đê - 0782657635 - lekhoide@gmail.com

Project Objectives : Thiết kế và xây dựng Website “Cổng thông tin sự kiện tại Đại học Công Nghệ - Hutech” trên nền tảng công nghệ ASP.NET MVC 5 cho Đò Án Tổng Hợp và sử dụng công nghệ Crawl Data để cào dữ liệu hướng dẫn bởi giảng viên Lê Thị Ngọc Thơ trong vòng 10 tuần.

Approach :

- Xác định đề tài, công nghệ, thời gian thực hiện 2 tuần đầu tiên. Phân tích và thiết kế đề tài trong 5 tuần. Thực hiện đề tài 2 tuần và tuần cuối nộp báo cáo và chuẩn bị Demo, bài thuyết trình.
- Mỗi tuần nộp báo cáo cho giáo viên hướng dẫn và nhận yêu cầu từ giảng viên hướng dẫn nếu có.

Roles and Responsibilities :

Role	Name	Organization / Position	Contact Information	Mssv / Class
Phân tích thiết kế	LÊ KHÔI ĐÊ	Design	0782657635 lekhoide@gmail.com	14110101 02 15DTH06
Viết code	NGUYỄN BÁ TÂM	Development	0333058893 tamnguyen.tgth@gmail.com	16110610 22 16DTHC 2

Sign-off: (Signatures of all above stakeholders. Can sign by their names in table above.)

LÊ KHÔI ĐÊ

NGUYỄN BÁ TÂM

1. 2 Cấu trúc phân rã công việc (WBS)

1. 2. 1 Xác định (10/02/2020 - 29/02/2020)

1. 2. 1. 1 Xác định ý tưởng thực hiện đề tài (10/02/2020 - 29/02/2020)

Kết quả: Giới thiệu về đề tài.docx

Chịu trách nhiệm: Lê Khôi Đệ.

1. 2. 1. 2 Xác định thực trạng vấn đề / mục đích lý do về ý tưởng thực hiện đề tài (10/02/2020 - 29/02/2020)

Kết quả: Thực trạng vấn đề.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 1. 3 Xác định yêu cầu về giao diện (24/02/2020 - 28/02/2020)

Kết quả: Yêu cầu giao diện.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 1. 4 Xác định công nghệ và ngôn ngữ lập trình sử dụng thực hiện đề tài (24/02/2020 - 28/02/2020)

Kết quả: Giới thiệu công nghệ.docx

Chịu trách nhiệm: Nguyễn Bá Tâm - Lê Khôi Đệ

1. 2. 1. 5 Xác định thời gian bắt đầu và kết thúc của đề tài

Thời gian thực hiện đồ án: 10/02/2020 - 6/06/2020

Thời gian nộp báo cáo và Poster đề tài: 8/06/2020 - 12/06/2020

Thời gian báo cáo trước hội đồng: 20/06/2020 - 27/06/2020

Chịu trách nhiệm: Khoa Công Nghệ Thông Tin

1. 2. 2 Phân tích và thiết kế (1/03/2020 - 04/04/2020)

1. 2. 2. 1 Phân tích chức năng (01/03/2020 - 04/04/2020)

- Crawl dữ liệu (01/03/2020 - 04/04/2020)

Kết quả: Tài liệu phân tích nghiên cứu 1.docx

Chịu trách nhiệm: Nguyễn Bá Tâm

- Dữ liệu khi cào về được Import vào Database(01/03/2020 - 04/04/2020)

Kết quả: Tài liệu phân tích nghiên cứu 2.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 2. 2 Phân tích - thiết kế giao diện (22/03/2020 - 04/04/2020)

- Giao diện trang chủ người dùng (22/03/2020 - 28/03/2020)
Kết quả: Tài liệu hình ảnh và kèm lời chú thích 1.docx
Chịu trách nhiệm: Lê Khôi Đệ
- Giao diện người dùng xem chi tiết một sự kiện (29/03/2020 - 04/04/2020)
Kết quả: Tài liệu hình ảnh và kèm lời chú thích 2.docx
Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 3 Thực hiện (05/04/2020 - 23/05/2020)

1. 2. 3. 1 Cập nhật thông tin lên hệ thống (05/04/2020 - 11/04/2020)

Cập nhật cơ sở dữ liệu Hutech 2020 trên SQL Server

Kết quả: Tài liệu hình ảnh và kèm lời giải thích.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 3. 2 Thực hiện chức năng (05/04/2020 - 23/05/2020)

- Crawl dữ liệu (05/04/2020 - 23/05/2020)

Đầu vào: Tài liệu phân tích nghiên cứu 1.docx

Kết quả: Tài liệu hình ảnh và kèm lời giải thích 1.docx

Chịu trách nhiệm: Nguyễn Bá Tâm

- Dữ liệu khi cào về được Import vào Database (03/05/2020 - 23/05/2020)

Đầu vào: Tài liệu phân tích nghiên cứu 2.docx

Kết quả: Tài liệu hình ảnh và kèm lời giải thích 2.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 3. 3 Thực hiện giao diện lên hệ thống (12/04/2020 - 02/05/2020)

Kết quả: Tài liệu hình ảnh và kèm lời giải thích.docx

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 4 Kiểm thử (24/05/2020 - 30/05/2020)

Kiểm thử chức năng (24/05/2020 - 30/05/2020)

Kết quả: Tài liệu đánh giá kết quả.docx

Chịu trách nhiệm: Nguyễn Bá Tâm - Lê Khôi Đệ

1. 2. 5 Kết luận (31/05/2020 - 27/06/2020)

1. 2. 5. 1 Viết và chỉnh sửa báo cáo(31/05/2020 - 12/06/2020)

Chịu trách nhiệm: Lê Khôi Đệ

1. 2. 5. 2 Làm Slide thuyết trình (13/06/2020 - 20/06/2020)

Chịu trách nhiệm: Nguyễn Bá Tâm - Lê Khôi Đệ

1. 2. 5. 3 Demo - thuyết trình (20/06/2020 - 27/06/2020)

Chịu trách nhiệm: Nguyễn Bá Tâm

1. 3 Cấu trúc đồ án

Trang phụ bìa

Lời cam đoan

Mục lục

Danh mục các bảng

Danh mục các hình vẽ

Chương 1 Tổng quan

Giới thiệu đề tài

Cấu trúc phân rã công việc

Cấu trúc đồ án

Chương 2 Cơ sở lý thuyết

Thực trạng vấn đề

Giới thiệu công nghệ

Chương 3 Phân tích và thiết kế

Đặc tả yêu cầu

Mô hình

Crawl Data

Giao diện

Chương 4 Cài đặt và đánh giá kết quả

Crawl Data

Nạp JSON lên SQL đầy lên hệ thống

Chức năng của hệ thống

Chương 5 Kết luận

Kết luận chung

Khẳng định những kết quả đạt được

Đánh giá phần mềm

Hướng phát triển và mở rộng đề tài

Tài liệu tham khảo

CHƯƠNG 2 CƠ SỞ LÝ THUYẾT

2. 1 Thực trạng vấn đề

Dự án “Xây dựng cổng thông tin sự kiện tại Đại Học Công Nghệ - Hutech” được xây dựng nhằm hỗ trợ sinh viên thấy được nhiều thông tin hơn, thông tin ở đây là các hội thảo, sự kiện, câu lạc bộ, đội nhóm được chia ra riêng lẻ ở 7 Trung Tâm, 8 Viện, 11 Phòng Ban, 13 Khoa và nhiều nhó lẻ khác mà giờ đây sẽ được tổng hợp qua một Website. Qua phân tích khảo sát 100 bạn sinh viên Hutech, thì nhóm em rút được ba vấn đề cơ bản sau :

1. Đa số sinh viên tham gia hội thảo, sự kiện, câu lạc bộ,... Thứ nhất để cộng điểm rèn luyện, thứ hai vì niềm yêu thích, tò mò. Nhưng hiện nay đa số các hội thảo, sự kiện được cập nhật trên Facebook mà ở từng trang khác nhau, nhóm khác nhau gây tình trạng dễ bị bỏ sót nếu không thường xuyên cập nhật
2. Trên Website trường Hutech đa số chỉ cập nhật các hội thảo, sự kiện nổi bật bỏ qua các hội thảo, sự kiện nhỏ lẻ khác và thường nằm trong từng danh mục không phù hợp với người dùng khó tính và thiếu kiên nhẫn.
3. Còn các hội thảo, sự kiện nhỏ lẻ khác thì chỉ được truyền miệng qua các lớp học, các giảng viên, băng rôn từng nơi trong trường, từng cơ sở.

2. 2 Giới thiệu công nghệ

Những công nghệ mà nhóm đã áp dụng vào đề tài :

- ✧ Công nghệ ASP.NET.MVC 5 của Microsoft

Những công cụ mà nhóm đã áp dụng vào đề tài :

- ✧ Công cụ Visual Studio 2019
- ✧ Công cụ Microsoft SQL Server Management Studio 2017
- ✧ Thư viện xử lý văn bản *html* HtmlAgilityPack và Fizzler
- ✧ Thư viện xử lý dữ liệu dạng Json: Newtonsoft. Json

Những ngôn ngữ mà nhóm đã áp dụng vào đề tài :

- ✧ Ngôn ngữ C#
- ✧ Ngôn ngữ HTML
- ✧ Ngôn ngữ CSS
- ✧ Ngôn ngữ JavaScript
- ✧ Bootstrap

2.2.1 Công nghệ ASP.NET MVC 5 của Microsoft là gì?

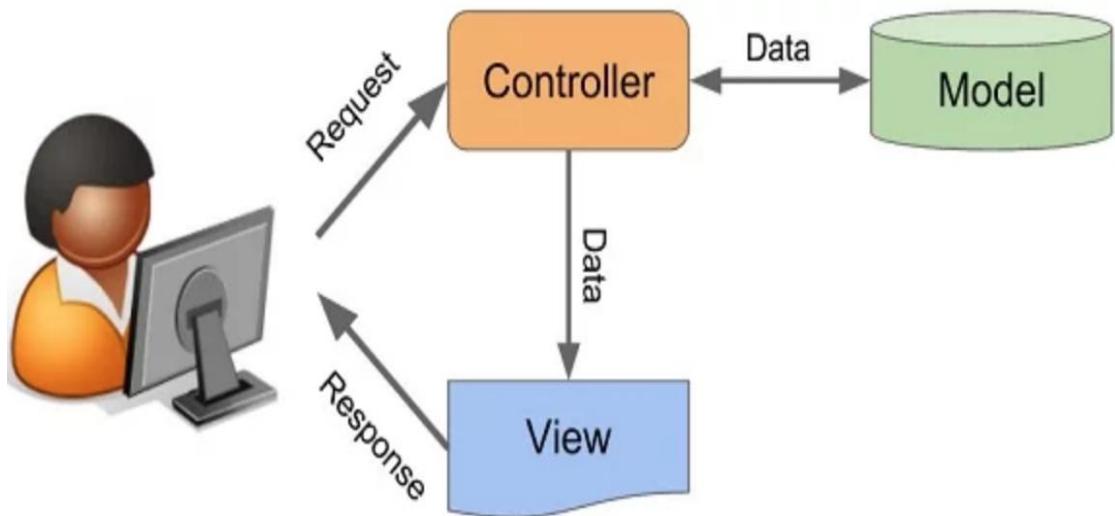
ASP.NET MVC 5 là một Framework lập trình web mới của Microsoft, công nghệ này ứng dụng mô hình MVC vào trong ASP.NET.

MVC là cụm từ viết tắt của Models - View - Controllers :

Models : nơi lưu trữ dữ liệu

View : giao diện người dùng được sử dụng bởi khách truy cập trang Web để nhìn thấy các dữ liệu

Controllers : chịu trách nhiệm xử lý các tương tác của người dùng với trang Web. Xác định loại View nào cần phải hiển thị và giao tiếp với Models.



Lợi ích ứng dụng Web dựa trên mô hình MVC

Dễ dàng quản lý sự phức tạp của ứng dụng bằng cách chia ứng dụng thành ba phần model, view, controller

Sử dụng mẫu Front Controller, mẫu này giúp quản lý các requests (yêu cầu) chỉ thông qua một Controller

Hỗ trợ tốt cho các ứng dụng được xây dựng bởi những đội có nhiều lập trình viên và thiết kế mà vẫn quản lý được tính năng của ứng dụng

2. 2. 2 Công nghệ SQL Server của Microsoft

SQL Server hay còn gọi là Microsoft SQL Server, viết tắt là MS SQL Server. Đây là một phần mềm được phát triển bởi Microsoft dùng để lưu trữ dữ liệu SQL Server cung cấp đầy đủ công cụ để quản lý, từ giao diện GUI cho đến việc sử dụng ngôn ngữ truy vấn SQL. Ngoài ra điểm mạnh của nó là Microsoft nên có khá nhiều nền tảng kết hợp hoàn hảo với SQL Server như ASP.NET, C# xây dựng Winform, bởi vì nó hoạt động hoàn toàn độc lập.



Ở SQL server có các tính năng giúp việc sử dụng hiệu quả hơn như:

- Cho phép tạo nhiều cơ sở dữ liệu
- Duy trì lưu trữ bền vững
- Bảo mật cao
- SQL Server hỗ trợ ngôn ngữ truy vấn SQL – ngôn ngữ tựa Tiếng Anh nên rất dễ sử dụng.

2. 2. 3 Khái niệm Crawl Data

Crawl data dữ liệu Web là một kỹ thuật thu thập dữ liệu từ các website trên mạng theo đường link cho trước. Các Web Crawler sẽ truy cập vào đường link và download toàn bộ dữ liệu cũng như tìm kiếm thêm các đường link bên trong để dowloand.

Nếu trong quá trình thu thập dữ liệu, ta chỉ chắt lọc những thông tin cần thiết phù hợp với nhu cầu thì đó gọi là web Scaping.

Để có thể Crawl được các dữ liệu trên trang web ta cần quan tâm đến các yếu tố như trang Web muốn crawl có bị chặn request hay không và cấu trúc của trang Web đó. Một trang Web có cấu trúc ổn định sẽ dễ dàng lấy dữ liệu hơn là trang Web có cấu trúc mỗi trang một định dạng khác nhau, bởi vì khi thực hiện kỹ thuật Crawl ta dựa chủ yếu vào các element để lấy dữ liệu.

CHƯƠNG 3 : PHÂN TÍCH VÀ THIẾT KẾ

3. 1 Đặc tả yêu cầu

Sự kiện của Website được chia thành 2 loại : sự kiện trong trường và sự kiện ngoài trường.

Các thông tin của sự kiện cần quản lý bao gồm : mã sự kiện, tên sự kiện, loại sự kiện, tên địa điểm, địa chỉ, phòng, thời gian bắt đầu, thời gian kết thúc, ngày bắt đầu, ngày kết thúc, thông tin sự kiện, ảnh sự kiện, ngày tạo sự kiện, tên cơ sở diễn ra, tên bộ phận tổ chức, tên quận/huyện diễn ra.

Một sự kiện được tổ chức bởi một hay nhiều bộ phận hay không bộ phận nào. Một bộ phận tổ chức được một hay nhiều sự kiện hay không sự kiện nào.

Một sự kiện được diễn ra ở một cơ sở. Một cơ sở sẽ diễn ra nhiều sự kiện.

Một sự kiện sẽ được diễn ra ở một quận huyện. Một quận huyện sẽ diễn ra nhiều sự kiện.

Admin là tác nhân duy nhất thực hiện được hành vi và bắt buộc phải đăng nhập để thực hiện được chức năng.

Dưới đây là một số mô tả về chức năng của Website.

3. 1. 1 Quản lý sự kiện

Gồm 4 chức năng chính: tạo sự kiện, xem sự kiện, sửa sự kiện, xóa sự kiện.

Xem sự kiện có chức năng tìm kiếm : tên sự kiện, loại sự kiện, tên cơ sở, tên bộ phận, tên quận huyện, thời gian (hôm nay, ngày mai, tuần này).

Xóa sự kiện có hành vi : xác nhận xóa.

Tạo sự kiện :

- + Mã sự kiện : thuộc tính bị ẩn. Được tự động tạo khi tạo sự kiện thành công.
Thuộc tính bắt buộc.
- + Tải ảnh bìa: kích thước tối ưu 1560 - 600px và dung lượng không lớn hơn 1 MB nếu không hợp lệ thì thông báo mời nhập lại. Thuộc tính không bắt buộc nhập.
- + Tên sự kiện: tối đa 128 ký tự nếu vượt quá thì thông báo mời nhập lại.
Thuộc tính bắt buộc nhập.
- + Loại sự kiện : chọn 1 trong 2 là “Trong trường” và “Ngoài trường”, nếu chưa chọn thì thông báo mời chọn. Thuộc tính bắt buộc chọn.
- + Tên địa điểm : tối đa 50 ký tự nếu vượt quá thì thông báo mời nhập lại.
Thuộc tính bị ẩn và chỉ hiện khi check ô “Địa chỉ tùy chọn” trên giao diện.
Thuộc tính không bắt buộc nhập.
- + Địa điểm : tối đa 128 ký tự nếu vượt quá thì thông báo mời nhập lại. Thuộc tính bị ẩn và chỉ hiện khi check ô “Địa chỉ tùy chọn” trên giao diện. Thuộc tính không bắt buộc nhập.
- + Tên quận huyện : chỉ được chọn một quận huyện. Thuộc tính bị ẩn và chỉ hiện khi check ô “Địa chỉ tùy chọn” trên giao diện. Thuộc tính không bắt buộc nhập.
- + Tên cơ sở : chỉ được chọn một cơ sở nếu không chọn thì thông báo mời chọn. Thuộc tính bắt buộc chọn.
- + Phòng : tối đa 20 ký tự nếu vượt quá thì thông báo mời nhập lại. Thuộc tính không bắt buộc nhập.
- + Tên bộ phận : chọn được một và nhiều hoặc không chọn bộ phận. Thuộc tính không bắt buộc chọn.

- + Thông tin sự kiện : Ràng buộc chỉ tối đa 4000 kí tự có thể chỉnh sửa, định dạng văn bản nếu vượt quá thì thông báo mời nhập lại. Thuộc tính không bắt buộc nhập
- + Ngày bắt đầu - ngày kết thúc : Hiện một danh sách lịch, tháng - năm trên đầu và danh sách ngày ở dưới. Những tháng - năm và ngày đã cũ thì sẽ bị mờ có ý nghĩa là không chọn được. Còn từ tháng - năm, ngày hiện tại và tương lai thì được chọn nhưng chỉ được chọn một cho ngày bắt đầu và ngày kết thúc. Ràng buộc ngày kết thúc chỉ được bằng và lớn hơn ngày bắt đầu nếu sai thông báo mời chọn lại. Định dạng hiển thị dd/mm/yyyy. Thuộc tính bắt buộc nhập
- + Thời gian bắt đầu - thời gian kết thúc : Hiện một danh sách giờ, hệ 12 giờ từ 12:00 AM đến 11:45 PM. Chỉ được chọn một cho thuộc tính thời gian bắt đầu và thuộc tính thời gian kết thúc. Ràng buộc giờ kết thúc phải lớn hơn giờ bắt đầu nếu sai thông báo mời chọn lại. Thuộc tính bắt buộc chọn
- + Ngày tạo sự kiện : thuộc tính bị ẩn trên giao diện. Tự động lấy ngày hiện tại khi tạo sự kiện thành công. Thuộc tính không bắt buộc

3. 1. 2 Quản lý bộ phận

- Gồm 4 chức năng : tạo bộ phận, xem bộ phận, sửa bộ phận, xóa bộ phận.
- Hutech hiện tại 2019 có 11 phòng ban, 13 khoa, 8 viện, 7 trung tâm.
- Các thông tin của bộ phận cần quản lý bao gồm : mã bộ phận, tên bộ phận, logo bộ phận.
- Admin phải đăng nhập để thực hiện chức năng này
- Xóa bộ phận có hành vi : xác nhận xóa.
- Tao bộ phận :
 - + Mã bộ phận : Thuộc tính bị ẩn trên giao diện. Được tự động tạo khi tạo bộ phận thành công. Thuộc tính bắt buộc
 - + Tên bộ phận : tối đa 128 kí tự nếu vượt quá thì thông báo nhập lại. Thuộc tính bắt buộc nhập
 - + Logo bộ phận : Kích thước tối ưu 130 - 130px nếu không hợp lệ thì mời nhập lại. Thuộc tính không bắt buộc nhập

3. 1. 3 Quản lý cơ sở

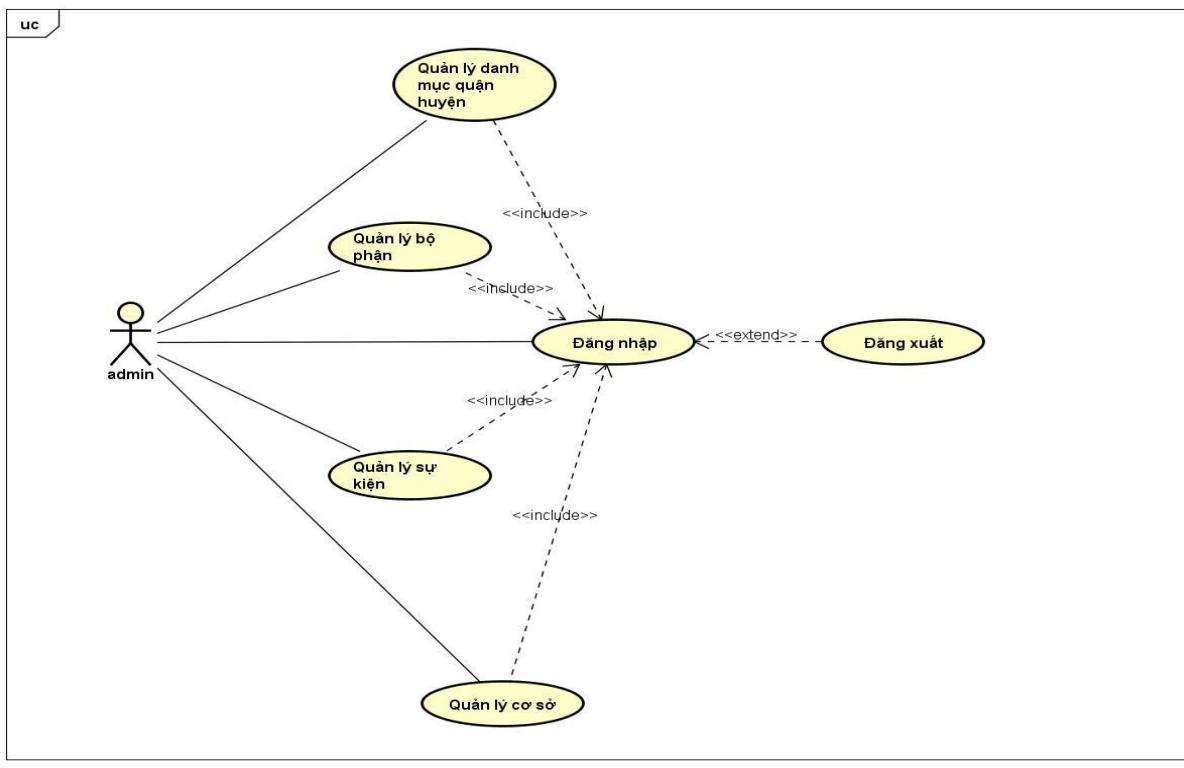
- Gồm 4 chức năng : tạo cơ sở, xem cơ sở, sửa cơ sở, xóa cơ sở
- Admin phải đăng nhập để thực hiện chức năng này.
- Hutech hiện tại 2019 có 5 cơ sở là :
 - A : 475A Điện Biên Phủ, P25, Q.Bình Thạnh, Tp.HCM
 - B : 475A Điện Biên Phủ, P25, Q.Bình Thạnh, Tp.HCM
 - U : 31/36 Ưng Văn Khiêm, P.25, Q.Bình Thạnh, Tp.HCM
 - E : Lô E1, Phân khu đào tạo E1, Khu Công Nghệ cao TP.HCM, Phường Hiệp Phú, Quận 9, TP.HCM.
 - R : Viện Công nghệ Cao Hutech, Lô E2B4, đường D1, Phường Long Thạnh Mỹ, khu Công Nghệ Cao, Quận 9, TP.HCM.
- Các thông tin của cơ sở cần quản lý bao gồm : mã cơ sở, tên cơ sở, địa chỉ.
- Xóa cơ sở có chức năng xác nhận xóa.
- Tạo cơ sở:
 - + Mã cơ sở : Thuộc tính bị ẩn trên giao diện. Được tự động tạo khi tạo cơ sở thành công. Thuộc tính bắt buộc
 - + Tên cơ sở : tối đa 20 kí tự nếu vượt quá thì thông báo lỗi nhập lại. Thuộc tính bắt buộc nhập
 - + Địa chỉ : tối đa 256 kí tự nếu vượt quá thì thông báo lỗi nhập lại. Thuộc tính không bắt buộc nhập
- 3. 1. 4 Admin đăng nhập*
- Các thông tin tài khoản cần quản lý bao gồm : tên đăng nhập, mật khẩu.
- Đăng nhập
 - + Tên đăng nhập : tối đa 20 kí tự nếu vượt quá thì thông báo lỗi nhập lại. Thuộc tính bắt buộc nhập.
 - + Mật khẩu : tối đa 50 kí tự nếu vượt quá thì thông báo lỗi nhập lại. Thuộc tính bắt buộc nhập.

3. 1. 5 Danh mục quận huyện

- Admin phải đăng nhập để thực hiện hành vi này.
- Gồm 4 chức năng chính : tạo quận huyện, xem quận huyện, sửa quận huyện, xóa quận huyện
- Các thông tin quận huyện bao gồm : mã quận huyện, tên quận huyện.
- Tạo quận huyện :
 - + Mã quận huyện : thuộc tính bị ẩn trên giao diện. Được tự động tạo khi tạo quận huyện thành công. Thuộc tính bắt buộc
 - + Tên quận huyện : tối đa 128 ký tự nếu vượt quá thì thông báo nhập lại. Thuộc tính bắt buộc nhập

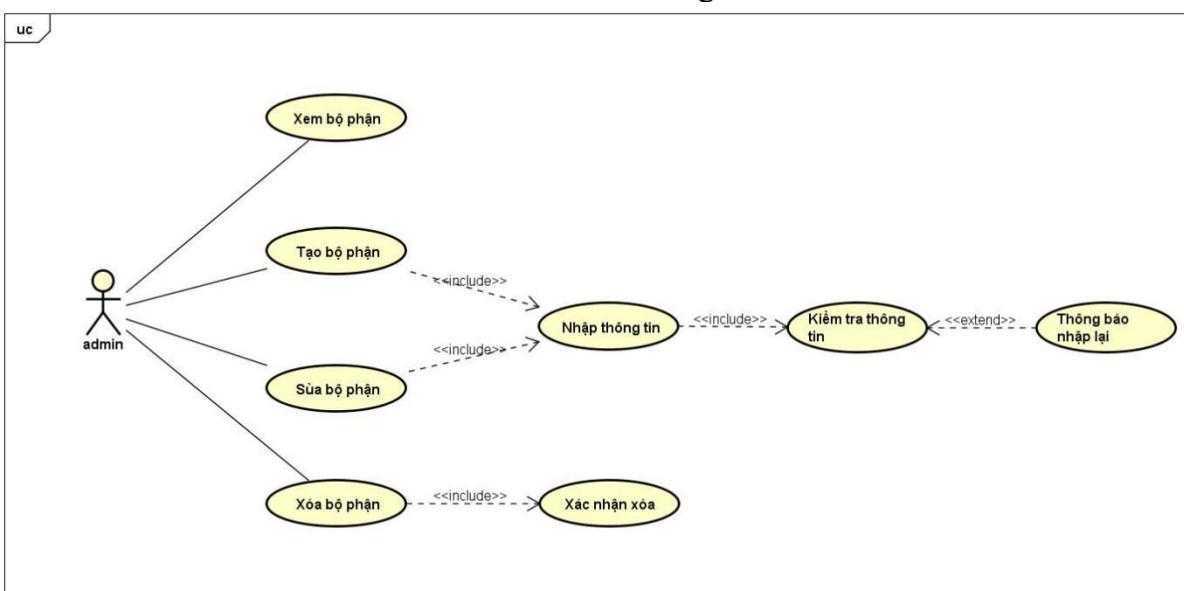
3.2 Mô hình

3.2.1 Lược đồ UseCase



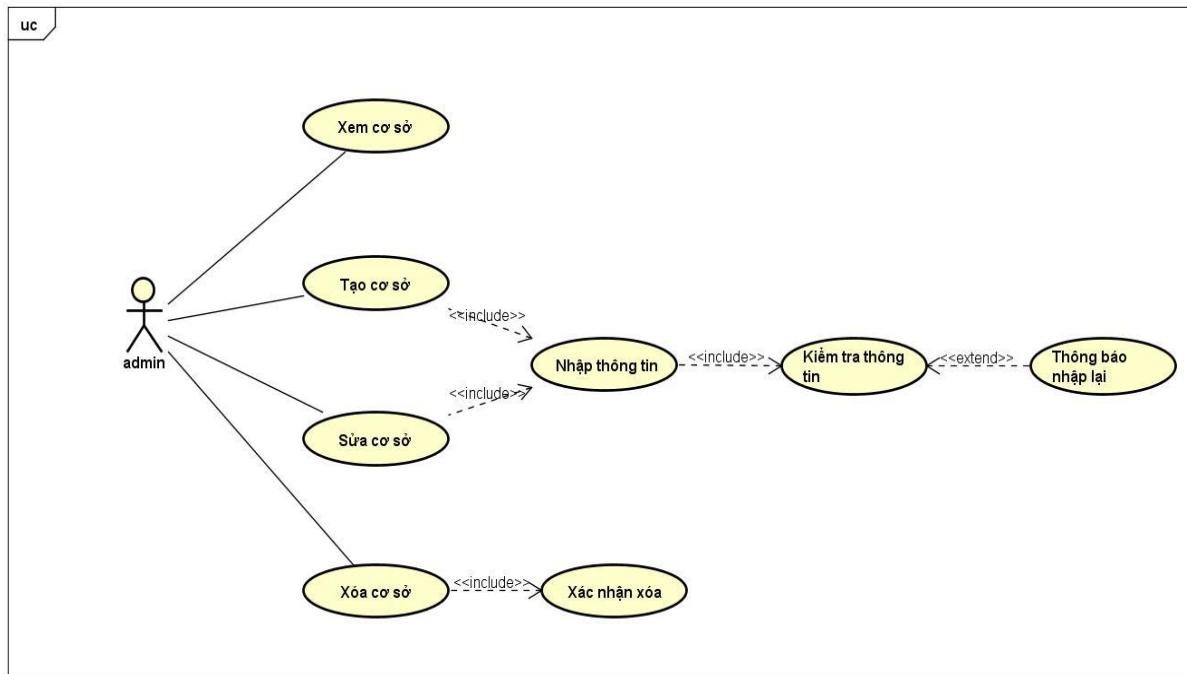
powered by Astah

Hình 3.1 UseCase tổng thể



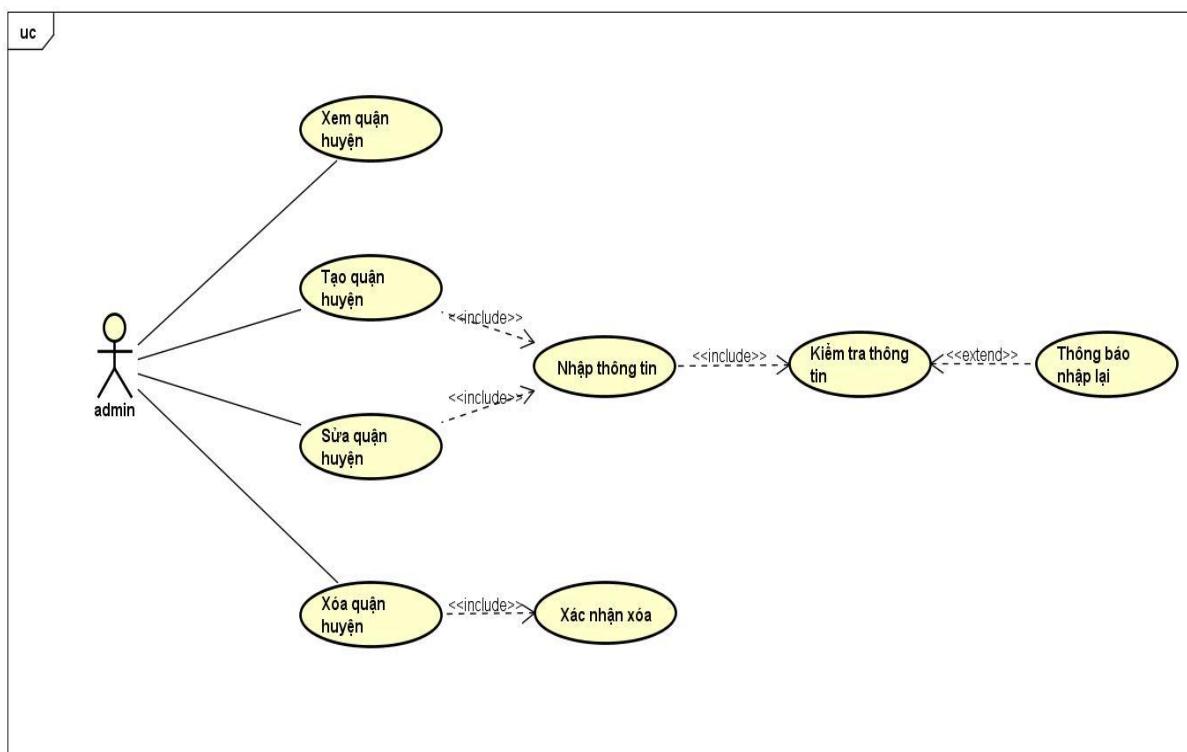
powered by Astah

Hình 3.2 UseCase quản lý bộ phận



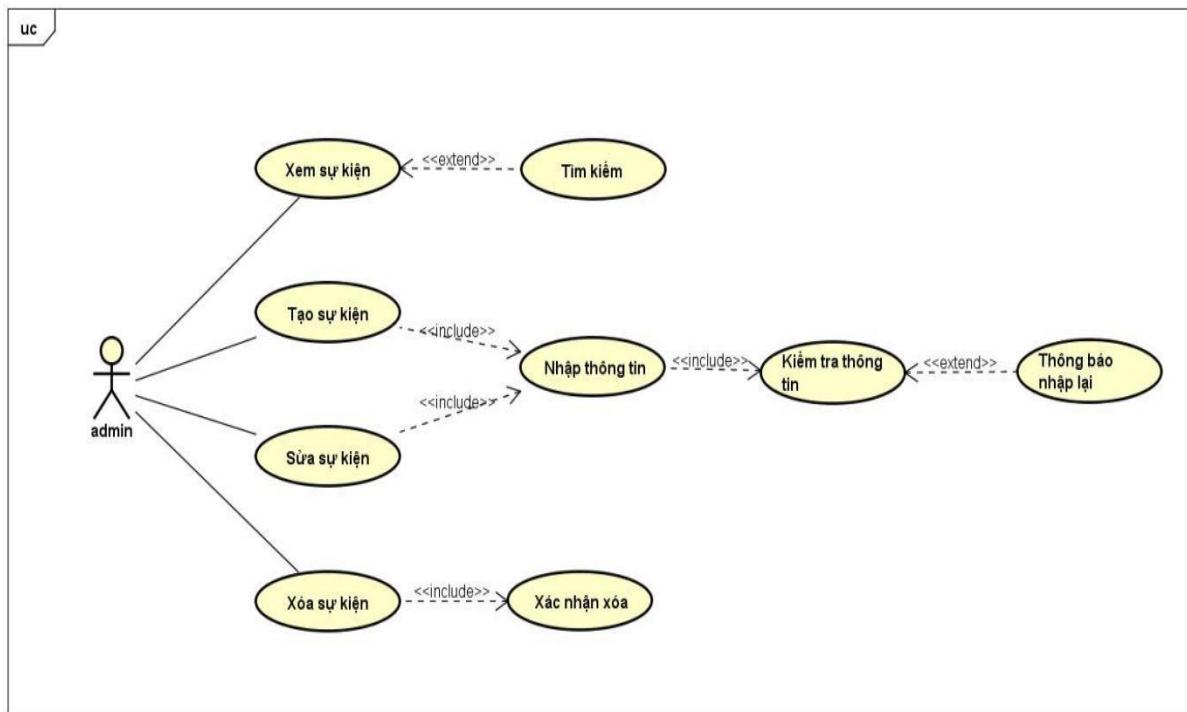
powered by Astah

Hình 3.3 UseCase quản lý cơ sở



powered by Astah

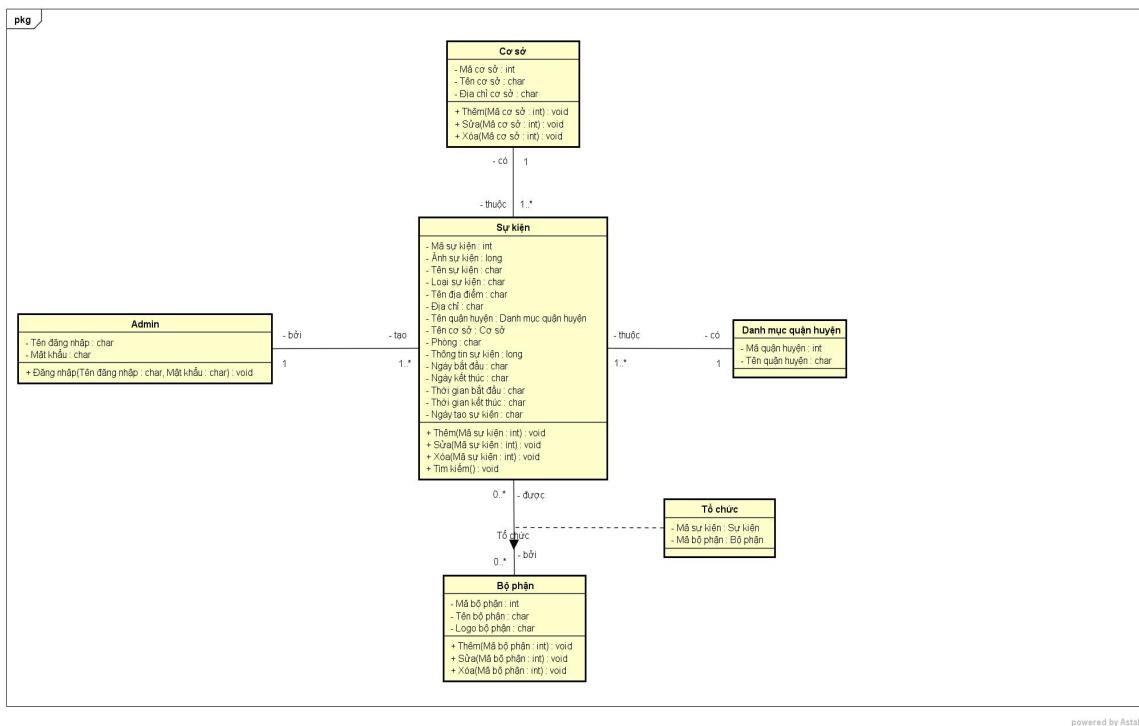
Hình 3.4 UseCase quản lý danh mục quận huyện



powered by Astah

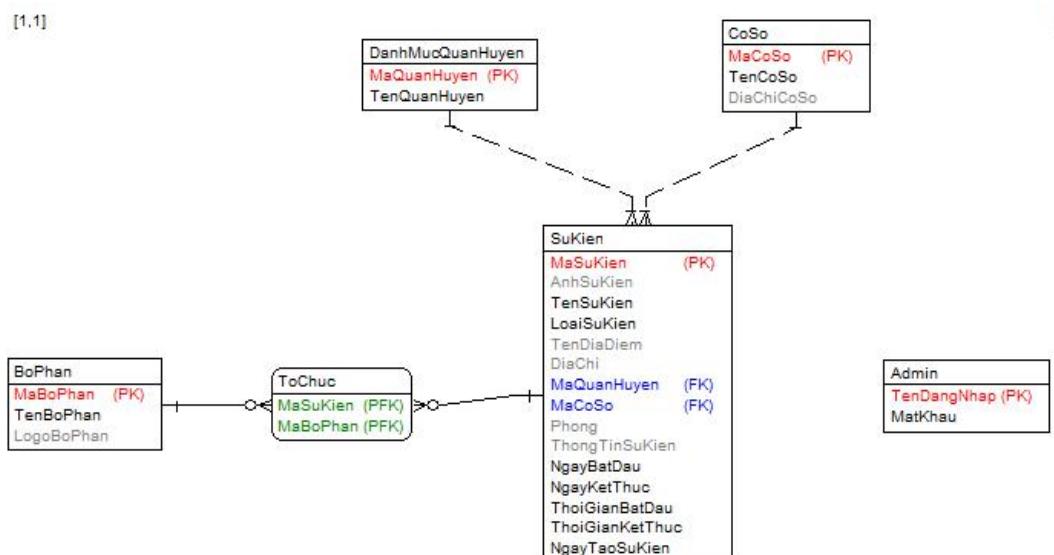
Hình 3.5 UseCase quản lý sự kiện

3.2.2 Class Diagram



Hình 3.6 Class Diagram tổng thể

3.2.3 ERD



Hình 3.7 ERD tổng thể

3.3 Crawl Data

3.3.1 Những thông tin cần trích xuất trên trang web

- Tên sự kiện
- Ảnh sự kiện
- Nội dung sự kiện
- Thời gian sự kiện(*nếu có*)

3.3.2 Phân tích cấu trúc trang web

3.3.2.1 Phân tích trang web khoa Dược trường Đại học công nghệ

Khi phân tích cấu trúc trang web của khoa Dược trường Đại học công nghệ ta nhận thấy:

- Mỗi thông tin sự kiện của trang Web tất cả đều được nằm trong một thẻ *div* có *class* = “*content-block*”, và tất cả các thông tin của từng sự kiện lại nằm trong thẻ *div* có tên *class* = “*item*”. Trong *class* “*item*” ta chú ý đến ba thông tin quan trọng là thẻ chứa tên của sự kiện, ảnh của sự kiện và nội dung của sự kiện.
- Trong đó tên của sự kiện nằm trong thẻ *a* có *class* = “*news - title*” là con của một thẻ *div* có *class* = “*col-sm-9 no-padding-left text-justify*” do những *class* này đều độc lập nên khi trích xuất tên của sự kiện ta chỉ cần lấy thông tin từ trong thẻ *a* có *class* = “*news - title*”.

```
▼<div class="content-block">
  ▶<div class="col-sm-12 no-padding-left padding-bottom">..</div>
  ▷<div class="item">
    ▷<div class="col-sm-3 no-padding-left">
      ▷<a href="https://www.hutech.edu.vn/khoaduoc/tin-tuc-su-kien/14583554-de-phong-dich-benh-covid-19-an-toan-sinh-vien-hutech-nghi-hoc-den-het-153">
        <img class="img-responsive" alt="https://file1.hutech.edu.vn/file/editor/homepage1/620-413%281%29.jpg" onerror="this.src=https://file1.hutech.edu.vn/file/news/default.jpg"
      </a>
    </div>
    ▷<div class="col-sm-9 no-padding-left text-justify">
      ▷<a href="https://www.hutech.edu.vn/khoaduoc/tin-tuc-su-kien/14583554-de-phong-dich-benh-covid-19-an-toan-sinh-vien-hutech-nghi-hoc-den-het-153" class="news-title">
        "Dé phòng dịch bệnh Covid-19 an toàn, sinh viên HUTECH nghỉ học đến hết 15/3" == $0
      </a>
    </div>
    ▷<span class="des_news">
      "Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...">
    </span>
  </div>
</div>
```

Hình 3.8 Phân tích Website định cao

TIN TỨC - SỰ KIỆN

Padding 5px 0px 0px
Contrast Aa 12.63 ✓



Để phòng dịch bệnh Covid-19 an toàn, sinh viên HUTECH nghỉ học đến hết 15/3

Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...

Hình 3.9 Phân tích Website định cào 1

- Ảnh của sự kiện là một đường dẫn đuôi *jpg* nằm trong thẻ *img* có *class=“img-responsive”*. Khi trích xuất hình ảnh từ trang Web ta chỉ cần truyền vào thẻ *img* lấy đường dẫn *src* thì sẽ lấy được thông tin hình ảnh của sự kiện đó.

```
▼<div class="col-sm-3 no-padding-left">
  <a href="https://www.hutech.edu.vn/khoaduoc/tin-tuc-su-kien/14583192-hutech-trang-trong-to-chuc-le-ky-niem-70-nam-ngay-truyen-thong-sinh-vien-hoc-sinh">
     == $0
  </a>
</div>
```



Để phòng dịch bệnh Covid-19 an toàn, sinh viên HUTECH nghỉ học đến hết 15/3

Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...

Hình 3.10 Phân tích Website định cào 2

- Nội dung của dự kiến được thể hiện ở tron thẻ *span* có *class = “des-news”* là thẻ con của thẻ *div* có *class = “col-sm-9 no-padding-left text-justify”*

```

▼<div class="col-sm-9 no-padding-left text-justify"> == $0
▶<a href="https://www.hutech.edu.vn/khoeduoc/tin-tuc-su-kien/14583554-de-phong-dich-benh-covid-19-an-toan-sinh-vien-hutech-nghi-hoc-den-het-153" class="news-title">...</a>
▼<span class="des_news">
  "
    Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...
  "
</span>
```

Hình 3.11 Phân tích Website định cào 2



Để phòng dịch bệnh Covid-19 an toàn, sinh viên HUTECH nghỉ học đến hết 15/3
Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...

Hình 3.12 Phân tích Website định cào 3

3. 3. 2. 2 Phân tích trang web khoa công nghệ thông tin trường đại học công nghệ

Tương tự như phân tích trang web của khoa Dược, khi phân tích trang Web của khoa Công nghệ thông tin trường Đại học công nghệ ta cũng thu được:

- Mỗi thông tin sự kiện của trang Web tất cả đều được nằm trong một thẻ *div* có *class* = “*content-block*”, và tất cả các thông tin của từng sự kiện lại nằm trong thẻ *div* có tên *class* = “*item*”. Trong *class* “*item*” ta chú ý đến ba thông tin quan trọng là thẻ chứa tên của sự kiện, ảnh của sự kiện và nội dung của sự kiện.



```
▼<div class="content-block"> == $0
  ▶<div class="col-sm-12 no-padding-left padding-bottom">...</div>
  ▶<div class="item">...</div>
  ▶<div class="paging">...</div>
</div>
</div>
```

Hình 3.13 Phân tích Website định cào 4

TIN TỨC SỰ KIỆN



Cùng sinh viên Công nghệ thông tin HUTECH tìm hiểu về Deep Learning

Sáng ngày 08/10, Hội thảo NVIDIA - Fundamentals of Deep Learning for Computer Vision dành riêng cho các bạn sinh viên đang theo học ngành Công nghệ thông tin HUTECH đã diễn ra tại Trụ sở chính trường Đại học Công nghệ...



"Đô thị sáng tạo - Thách thức và triển vọng cho sinh viên" ngành Công nghệ thông tin HUTECH

Đó cũng chính là chủ đề của chương trình Kết nối doanh nghiệp với sinh viên diễn ra vào chiều ngày 18/10, do trường Đại học Công nghệ TP.HCM (HUTECH) phối hợp với Hội Tin học TP.HCM (HCA) tổ chức. Chương...

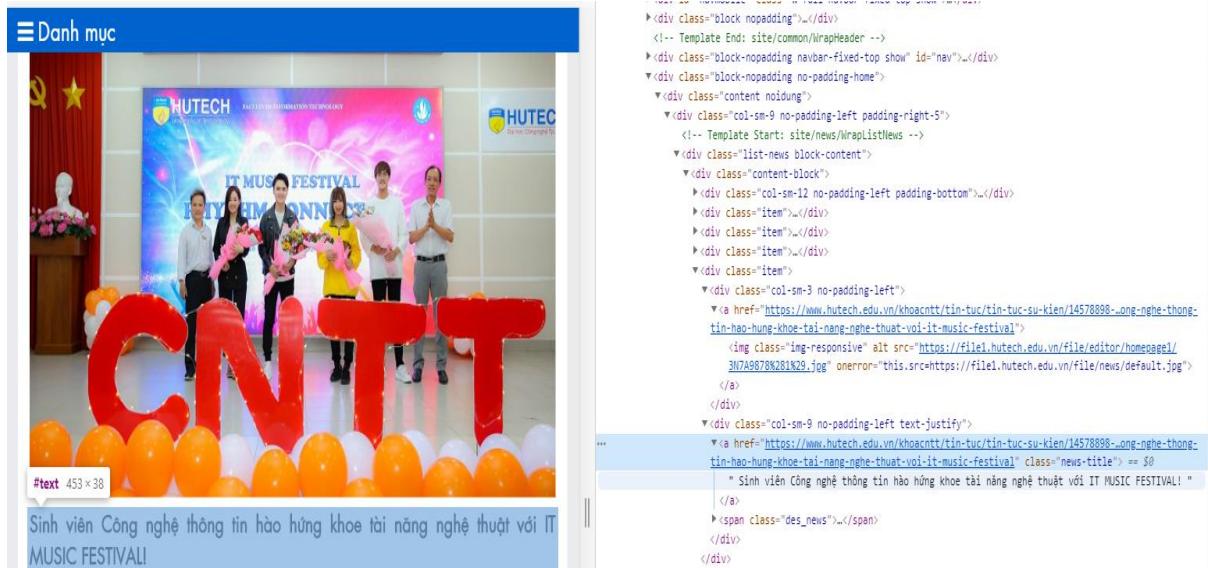


Những tài năng lập trình mới của khoa Công nghệ thông tin đã "lộ diện"!

Chiều 19/10, khoa Công nghệ thông tin trường Đại học Công nghệ TP.HCM (HUTECH) tổ chức cuộc thi "Ươm mầm tài năng lập trình HUTECH" - sân chơi học thuật thường niên dành cho sinh viên năm nhất nhóm ngành Công nghệ...

Hình 3.14 Phân tích Website định cào 5

- Tên của sự kiện nằm trong thẻ *a* có *class* = “*news - title*” là con của một thẻ *div* có *class* = “*col-sm-9 no-padding-left text-justify*” do những *class* này đều độc lập nên khi trích xuất tên của sự kiện ta chỉ cần lấy thông tin từ trong thẻ *a* có *class* = “*news - title*”.



Hình 3.15 Phân tích Website định cào 6

- Nội dung của dự kiện được thể hiện ở tron thẻ *span* có *class* = “*des-news*” là thẻ con của thẻ *div* có *class* = “*col-sm-9 no-padding-left text-justify*”.



Hình 3.16 Phân tích Website đinh cào 7

- Ảnh của sự kiện là một đường dẫn đuôi *jpg* nằm trong thẻ *img* có *class=“img-responsive”*. Khi trích xuất hình ảnh từ trang Web ta chỉ cần truyền vào thẻ *img* lấy đường dẫn *src* thì sẽ lấy được thông tin hình ảnh của sự kiện đó.

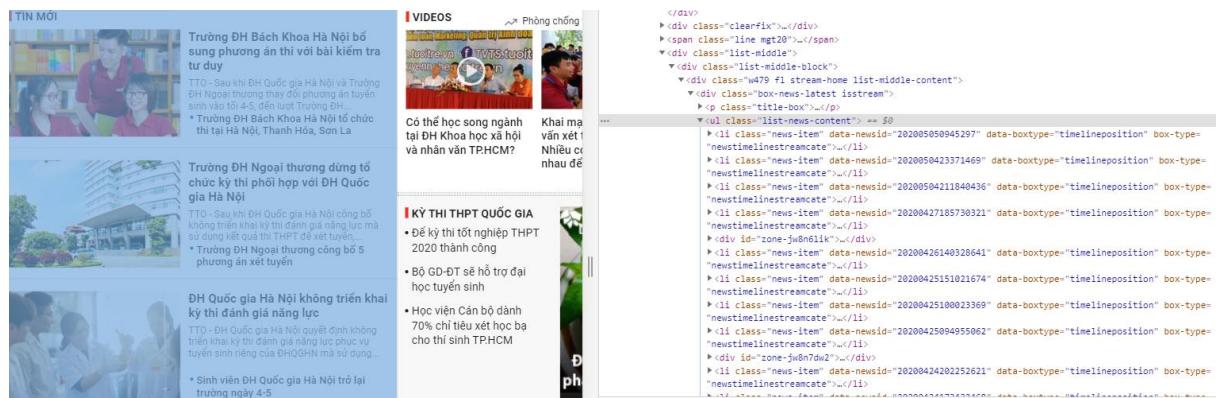


Hình 3.17 Phân tích Website định cào 8

3. 3. 2. 3 Phân tích Website ngoài trường Đại học công nghệ

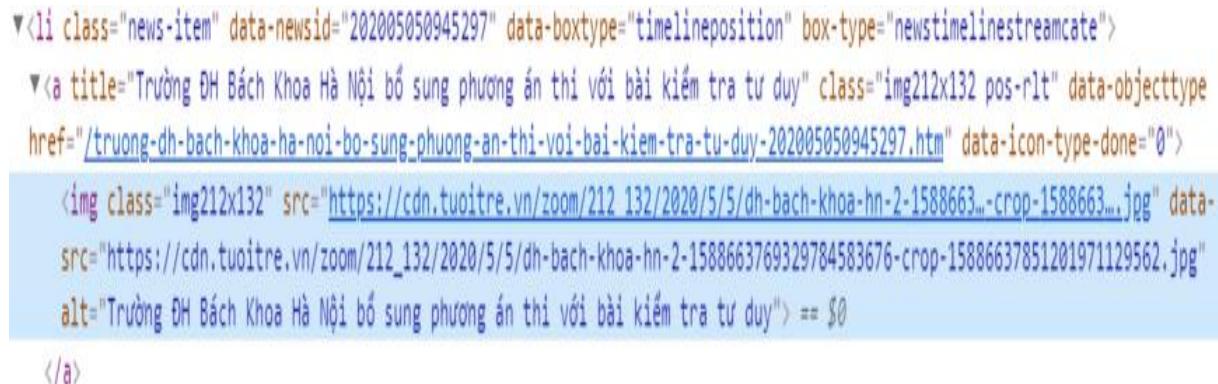
Ở đây ta đi vào phân tích trang web *Tuoitre.VN* và ta thấy được:

- Tất cả các sự kiện của trang web đều nằm trong thẻ *ul* với tên *class=“list-news-content”*, mỗi sự kiện riêng biệt nằm trong thẻ *li* là thẻ con của thẻ *ul* với *class = “news-item”*.



Hình 3.18 Phân tích Website định cào 9

- Phân tích vào sâu hơn thì ta có được hình ảnh của mỗi sự kiện nằm trong thẻ *img* có *class = “img212x132”*.



Hình 3.19 Phân tích Website định cào 10



Trường ĐH Bách Khoa Hà Nội bổ sung phương án thi với bài kiểm tra tư duy

TTO - Sau khi ĐH Quốc gia Hà Nội và Trường ĐH Ngoại thương thay đổi phương án tuyển sinh vào tối 4-5, đến lượt Trường ĐH...

- **Trường ĐH Bách Khoa Hà Nội tổ chức thi tại Hà Nội, Thanh Hóa, Sơn La**

Hình 3.20 Phân tích Website định cào 11

- Tên của các sự kiện đều nằm trong thẻ *h3* có *class* = “*title-news*”

```
<a title="Trường ĐH Bách Khoa Hà Nội bổ sung phương án thi với bài kiểm tra tư duy" class="img212x132 pos-rlt" data-objecttype href="/truong-dh-bach-khoa-ha-noi-bo-sung-phuong-an-thi-voi-bai-kiem-tra-tu-duy-202005050945297.htm" data-icon-type-done="0"></a>
<div class="name-news">
<h3 class="title-news">
<a title="Trường ĐH Bách Khoa Hà Nội bổ sung phương án thi với bài kiểm tra tư duy" data-objecttype data-id="202005050945297" data-linktype="newsdetail" data-comment="202005050945297" href="/truong-dh-bach-khoa-ha-noi-bo-sung-phuong-an-thi-voi-bai-kiem-tra-tu-duy-202005050945297.htm">Trường ĐH Bách Khoa Hà Nội bổ sung phương án thi với bài kiểm tra tư duy</a>

```

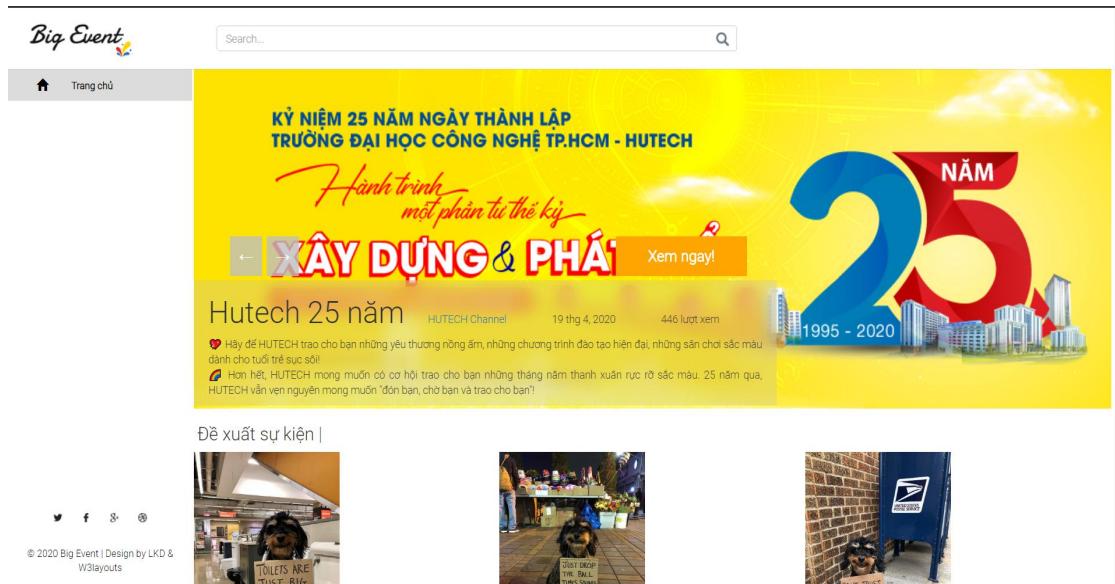
Hình 3.21 Phân tích Website định cào 12

3.3 Giao diện

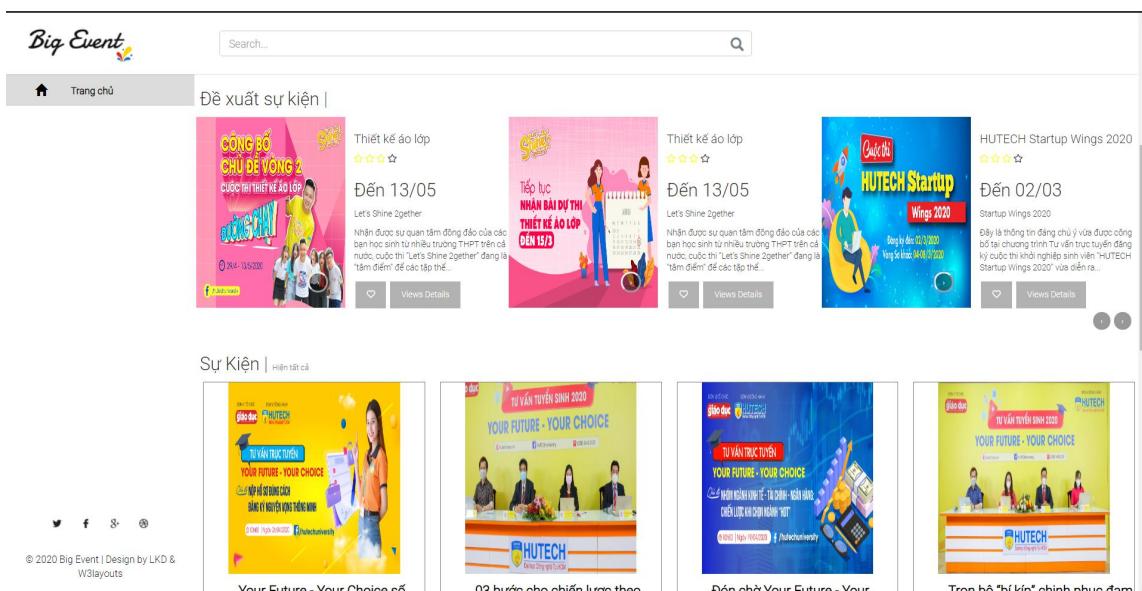
Yêu cầu tổng thể giao diện người dùng:

- ✧ Điều hướng đơn giản, người dùng phải tìm thấy nội dung trong vài giây.
- ✧ Bố cục đơn giản, hình ảnh sự kiện được đóng khung làm nổi bật hình ảnh bên trong, các khung có trúc gọn gàng xung quanh tạo cảm giác sắp xếp có thứ tự.
- ✧ Logo tránh mờ mép, hình ảnh sự kiện rõ ràng - tránh bị nhòe.
- ✧ Tập trung vào trải nghiệm người dùng.

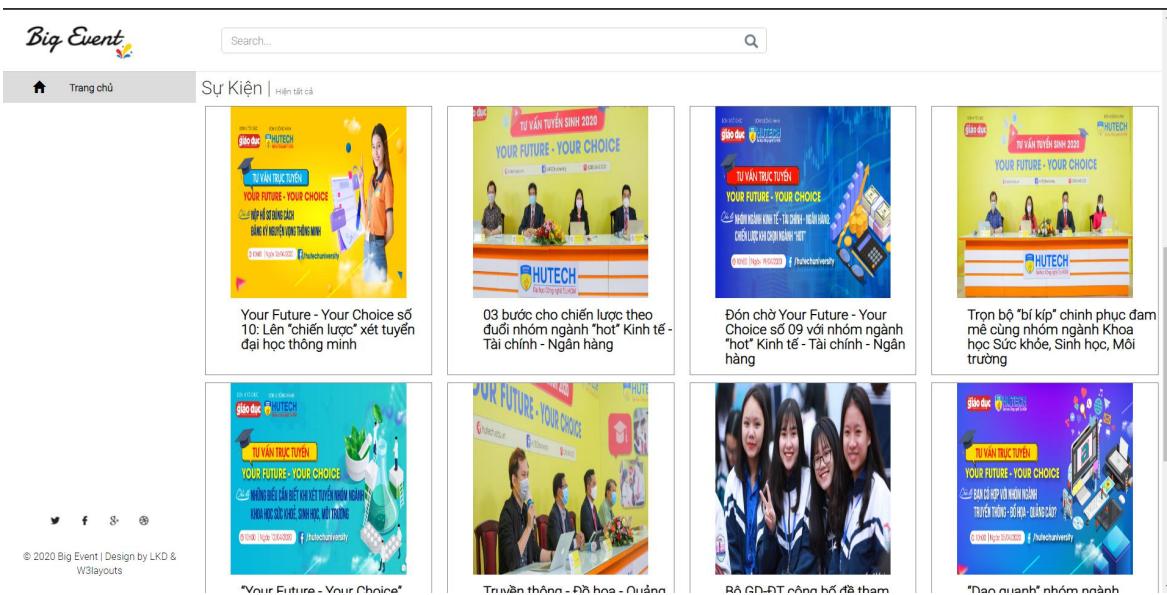
3.3.1 Giao diện trang chủ



Hình 3.22 Giao diện trang chủ người dùng



Hình 3.23 Giao diện trang chủ người dùng 1



Hình 3.24 Giao diện trang chủ người dùng 2

3. 3. 2 Giao diện xem chi tiết sự kiện

Hình 3.25 Giao diện xem chi tiết sự kiện người dùng

Ở dưới ảnh là các thông tin liên quan đến sự kiện đó

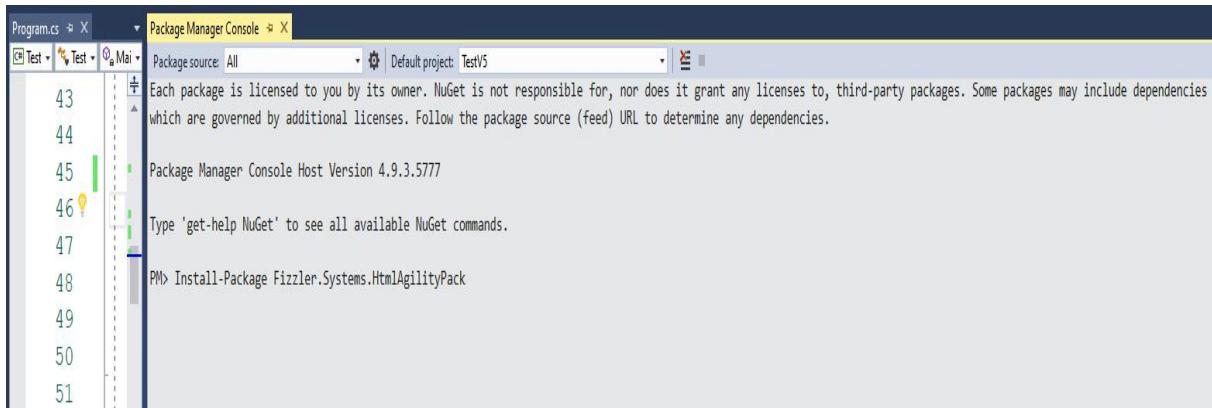
Ở bên phải trang là thanh các sự kiện liên quan

CHƯƠNG 4: CÀI ĐẶT VÀ ĐÁNH GIÁ KẾT QUẢ

4.1 Crawl Data

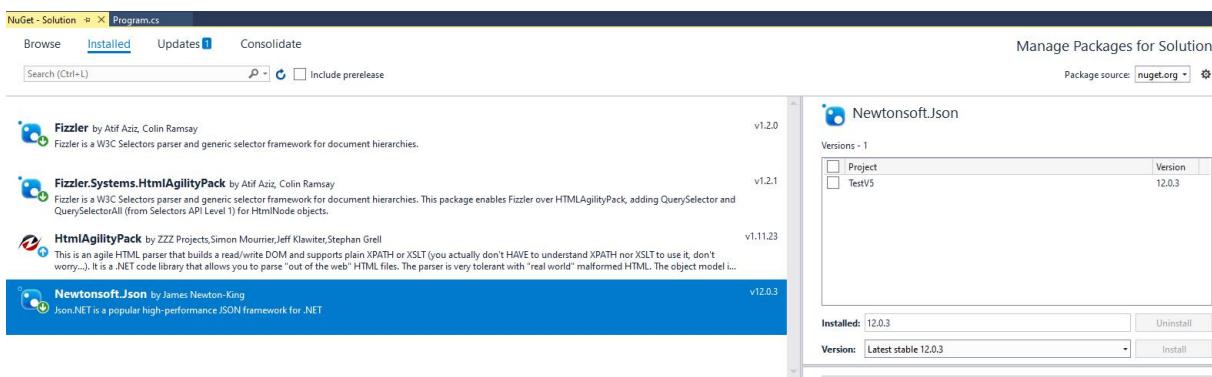
4. 1. 1 Cài đặt thư viện

- Sau khi tạo với một project ta vào Tool -> Package Manager Console gõ dòng lệnh *Install-Package Fizzler.Systems.HtmlAgilityPack* để cài đặt bộ thư viện Html Agility Pack.



Hình 4.26 Cài đặt môi trường Crawl Data

- Sau đó ta vào vào Manage Nuget Package For Sulution để cài đặt bộ thư viện **Newtonsoft.Json** hỗ trợ lưu trữ dữ liệu dưới dạng Json.



Hình 4.27 Cài đặt môi trường Crawl Data 1

4. 1. 2 Chương trình Crawl Data

- Mỗi chương trình là một đường dẫn đến trang Web để crawl dữ liệu về và lưu dưới dạng json.

```
HtmlWeb htmlWeb2 = new HtmlWeb()
{
    //AutoDetectEncoding = false,
    //OverrideEncoding = Encoding.UTF8
};
Console.OutputEncoding = Encoding.UTF8;//định dạng kiểu chữ và in ra
màn hình console

        //Load trang html của trang web cần crawl dữ liệu thông qua đường link
cho sẵn
        HtmlDocument          document2           =
htmlWeb2.Load("https://www.hutech.edu.vn/khoaduoc/tin-tuc/tin-tuc-su-kien");

        //load các thẻ....trong thẻ...
        var                  threadItems2        =
document2.DocumentNode.QuerySelectorAll("div.content-block > div.item").ToList();
        var items2 = new List<object>();
        foreach (var item in threadItems2)
        {
            //trích xuất các thẻ con của thẻ ...
            var linkNode = item.QuerySelector("a.news-title");//node lấy tên
của sự kiện
            var TenSuKien = linkNode.InnerText;

            var linkNode2 = item.QuerySelector("img.img-responsive");//node
lấy đường dẫn ảnh của sự kiện
            var AnhSuKien = linkNode2.Attributes["src"].Value;

            var linkNode3 = item.QuerySelector("span.des_news");
            var NoiDungSuKien = linkNode3.InnerText;
            items2.Add(new { TenSuKien ,AnhSuKien,NoiDungSuKien});

        }
        using          (StreamWriter          streamWriter        =
File.CreateText(@"D:\json2.json"))//lưu dưới dạng file json với tên là json2
        {
            JsonSerializer jsonSerializer = new JsonSerializer();
            jsonSerializer.Serialize(streamWriter, items2);
        }
        string      json2       =      JsonConvert.SerializeObject(items2,
Formatting.Indented);
        Console.WriteLine(json2);
        Console.ReadKey();
```

(Chương trình crawl dữ liệu trên trang web Khoa Dược trường đại học Công nghệ)

- Sau khi chạy chương trình thì ta nhận được kết quả như sau gồm tên sự kiện, đường dẫn ảnh sự kiện và nội dung sự kiện.

```

D:\ĐÁO ÁN TỔNG HỢP\Text5\Text5\bin\Debug\Text5.exe
{
    "TenSukien": "\n          Đã phòng dịch bệnh Covid-19 an toàn, sinh viên HUTECH nghỉ học đến hết 15/3\n          ",\n          "AnhSukien": "https://file1.hutech.edu.vn/file/editor/homepage/1/3NA6A439X281X29.png",\n          "NoiDungSukien": "Nhằm đảm bảo an toàn sức khỏe cho sinh viên - học viên trước tình hình diễn biến phức tạp của dịch bệnh Covid-19, trường Đại học Công nghệ TP.HCM (HUTECH) thông báo cho sinh viên - học viên tiếp tục nghỉ học đến hết...\n        ",\n        {
            "TenSukien": "\n          HUTECH trang trọng tổ chức Lễ kỷ niệm 70 năm ngày truyền thống Sinh viên, Học sinh\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/editor/homepage/1/3NA6A439X281X29.jpg",\n            "NoiDungSukien": "\n          Chiều 09/01/2020, đúng vào dịp kỷ niệm 70 năm ngày truyền thống Sinh viên, Học sinh (09/01/1950 - 09/01/2020), trường Đại học Công nghệ TP.HCM (HUTECH) đã long trọng tổ chức lễ kỉ niệm\n          70 năm ngày...\n        ",\n        {
            "TenSukien": "\n          HUTECH tổ chức Lễ kỷ niệm phản hồi của sinh viên về hoạt động giảng dạy Học kỳ I B năm học 2019 - 2020\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/editor/homepage/1/3A8P52.jpg",\n            "NoiDungSukien": "\n          Vào ngày 6/1/2020, trường Đại học Công nghệ TP.HCM (HUTECH) sẽ tổ chức khảo sát lấy ý kiến phản hồi của sinh viên về hoạt động giảng dạy của giảng viên nhằm nâng cao chất lượng dạy và học\n          c của Nhà trường.\n        ",\n        {
            "TenSukien": "\n          HUTECH lần đầu cả nước về số lượng cá nhân, tập thể \"Sinh viên 5 tốt\" cấp Trung ương\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/editor/homepage/1/626X2X284X29.png",\n            "NoiDungSukien": "\n          Vào ngày 6/1/2020, trường Đại học Công nghệ TP.HCM (HUTECH) sẽ tổ chức khảo sát lấy ý kiến phản hồi của sinh viên về hoạt động giảng dạy của giảng viên nhằm nâng cao chất lượng dạy và học\n          8/260 cá nhân (chiếm 33,5%) và...\n        ",\n        {
            "TenSukien": "\n          Đề án tuyển sinh Đại học chính quy năm 2020 dự kiến của HUTECH\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/editor/stories/hinh212-sea/IMG_5211.JPG",\n            "NoiDungSukien": "\n          Năm 2020, trường Đại học Công nghệ TP.HCM (HUTECH) dự kiến tuyển sinh 5.950 chỉ tiêu với&nbsp;47&nbsp;ngành đào tạo&nbsp;trình độ Đại học chính quy theo 04&nbsp;phương thức xét tuyển\n          đợt 1 và đợt 2.\n          Phương thức...\n        ",\n        {
            "TenSukien": "\n          Sinh viên Được HUTECH đón đợi Lễ tốt nghiệp đợt Tháng 12/2019\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/editor/homepage/1/ky-620.png",\n            "NoiDungSukien": "\n          Lễ tốt nghiệp cuối cùng trong năm 2019, từ ngày 07 đến ngày 15/12/2019 tới đây, trường Đại học Công nghệ TP.HCM (HUTECH) sẽ tổ chức Lễ trao bằng tốt nghiệp cho các Tân cử nhân, Tân\n          kỹ sư thuộc 14 Khoa/Viện...\n        ",\n        {
            "TenSukien": "\n          TB Tập huấn ban cán sự lớp học kỳ 2 năm học 2018-2019\n          ",\n            "AnhSukien": "https://imgnews/homepage/stories/hinh213-hea/ban-can-su20.png",\n            "NoiDungSukien": "\n          Nhằm rà soát Ban cán sự lớp trong toàn trường, triển khai hướng dẫn tổ chức Sinh hoạt lớp, Lớp học tiên tiến, công tác Dánh giá rèn luyện, đăng ký ngoại trú và các hoạt động, phong trào\n          Sinh viên của Đoàn Thành niên...\n        ",\n        {
            "TenSukien": "\n          Chương trình hỗ trợ vui xuân khoe - \"TẤY GIUN ĐẦU NĂM - SỐNG KHỎE QUANH NĂM\"\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/news/untitled-1546959233.jpg",\n            "NoiDungSukien": "\n          Việt Nam là một trong những quốc gia có điều kiện khí hậu thuận lợi cho bệnh giun tồn tại và phát triển. Theo điều tra của các Viện sét rết - Ký sinh trùng - Côn Trùng và các tỉnh thành\n          từ năm 2013-2017, tỷ lệ nhiễm...\n        ",\n        {
            "TenSukien": "\n          Khóa Dựp trao giải Pasteur cho Dựpt si tưng lai trong chương trình chào đón Tân sinh viên\n          ",\n            "AnhSukien": "https://imgnews/homepage/stories/hinh212-sea/IMG5008.jpg",\n            "NoiDungSukien": "\n          Chiều qua 19/09, hơn 400 Tân sinh viên khoa Dựpt, trường Đại học Công nghệ TP.HCM (HUTECH) đã có buổi giao lưu, gặp gỡ chính thức với Ban chủ nhiệm khoa tại Trụ sở chính của trường trên\n          Không khí vui tươi, cởi mở...\n        ",\n        {
            "TenSukien": "\n          Điều kiện mở nhà thuốc tư nhân: Dựpt si Đại học chỉ cần 18 tháng kinh nghiệm\n          ",\n            "AnhSukien": "https://file1.hutech.edu.vn/file/news/Img_0689_1-3492045657.jpg",\n            "NoiDungSukien": "\n          Tại Công ty vận trực tuyến của trường Đại học Công nghệ, cùng với việc tìm hiểu về thông tin tuyển sinh, chương trình đào tạo, cơ hội việc làm của...\n        "
    }
}

```

Hình 4.28 Kết quả Crawl Data

4. 1. 3 Dánh giá kết quả

- Qua quá trình phân tích trang web và thực hiện chương trình crawl dữ liệu lưu dưới dạng json ta nhận được kết quả mong muốn là tên của sự kiện, đường dẫn ảnh của mỗi sự kiện và nội dung của sự kiện đó.
- Một số trang web của các khoa trong trường không thể trích xuất dữ liệu, dữ liệu crawl về còn một số lỗi như xuất hiện nhiều khoảng trắng không mong muốn trong mỗi dữ liệu.

4.2 Nạp JSON vào SQL đẩy lên hệ thống

```

BigEvent7.sql - LKD\lekhoidc (62)* | JsonToSql.sql - LKD\lekhoidc (61)
File Edit View Query Project Tools Window Help
Object Explorer
BigEvent
New Query Execute
Nhập dữ liệu bảng tổ chức
INSERT INTO ToChuc(MaSuKien,MaBoPhan)
VALUES(1,2)
GO
INSERT INTO ToChuc(MaSuKien,MaBoPhan)
VALUES(1,3)
GO
INSERT INTO ToChuc(MaSuKien,MaBoPhan)
VALUES(2,2)
GO
INSERT INTO ToChuc(MaSuKien,MaBoPhan)
VALUES(2,3)
GO

--Nạp dữ liệu JSON lên hệ thống
DECLARE @json Nvarchar(max);
SELECT @json = BulkColumn FROM OPENROWSET(BULK 'D:\path.json' , SINGLE_CLOB) as j;
INSERT INTO [SuKien]([AnhSuKien], [TenSuKien])
SELECT AnhSuKien, TenSuKien
FROM OPENJSON(@json)
WITH([AnhSuKien] Nvarchar(max), [TenSuKien] Nvarchar(256))

SELECT * FROM SuKien
SELECT * FROM CoSo
SELECT * FROM DanhMucQuanHuyen
SELECT * FROM BoPhan
SELECT * FROM Admin
SELECT * FROM ToChuc

```

Messages

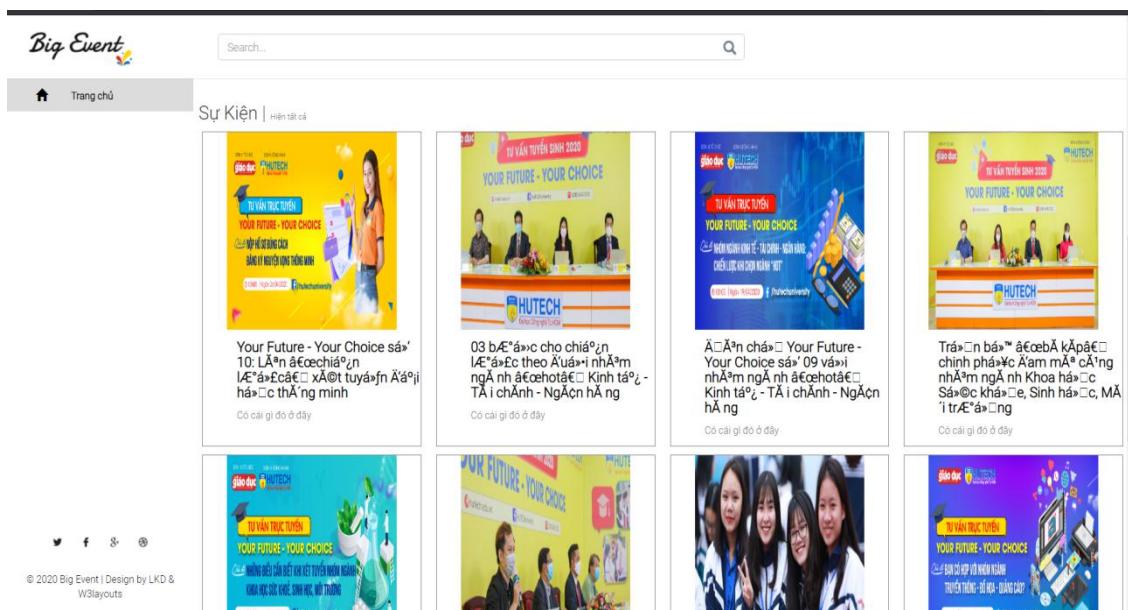
Query executed successfully.

Ready

Ln 345 Col 59 Ch 59 INS

Hình 4.29 Câu lệnh SQL đưa dữ liệu từ JSON vào hệ thống

Câu lệnh SQL đẩy file JSON lên hệ thống, chi tiết thì là đẩy vào bảng sự kiện với 2 thuộc tính là : Ảnh sự kiện, Tên sự kiện. Nhưng sẽ bị lỗi font chữ Việt ở thuộc tính Tên sự kiện



Hình 4.30 Dữ liệu từ JSON đã lên hệ thống

Cách giải quyết là vô hiệu hóa câu lệnh nạp file JSON thay vào đó là sao chép toàn bộ dữ liệu từ JSON vào file SQL định dạng lại.

```

GO
--Nap dữ liệu JSON lên hệ thống
DECLARE @json NVARCHAR(MAX) = N'[
    {
        "TenSuKien": "\nYour Future - Your Choice số 10: Lên \"chiến lược\" xét tuyển đại học thông minh\n",
        "NoiDungSuKien": "\nSau last chủ đề tập trung vào các nhóm ngành cụ thể, chương trình Tư vấn tuyển sinh trực tuyến Your Future - Your Choice số 10 sẽ được phát sóng.\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583886-young-future-your-choice-so-10-len-chien-luoc-xet-tuyen-dai-hoc-thong-minh",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/4eab-ff9d0878b85699a3e22651a231.png"
    },
    {
        "TenSuKien": "\n          03 bước cho chiến lược theo đuổi nhóm ngành \"hot\" Kinh tế - Tài chính - Ngân hàng\n",
        "NoiDungSuKien": "\n          Chương trình tư vấn trực tuyến Your Future - Your Choice số 19/4 vừa qua đã mang đến cho các bạn học sinh yêu thích nhóm ngành Kinh tế - Tài chính\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583882-03-buoc-cho-chien-luoc-theo-duoi-nhom-nhanh-hot-kinh-te-tai-chinh-ngan-hang",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/75c4e8ffa87ca373b95265e7b49a2b7.jpg"
    },
    {
        "TenSuKien": "\n          Bốn chờ Your Future - Your Choice số 09 với nhóm ngành \"hot\" Kinh tế - Tài chính - Ngân hàng\n",
        "NoiDungSuKien": "\n          Với chủ đề Nhóm ngành Kinh tế - Tài chính - Ngân hàng: Chiến lược khi chọn ngành \"hot\", chương trình Tư vấn trực tuyến hứa hẹn sẽ hỗ trợ giải đáp\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583858-don-chon-your-future-your-choice-so-09-voi-nhom-nhanh-hot-kinh-te-tai-chinh-ngan-hang",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/3b04uiddf1fa7faccfesfb03cb3879.png"
    },
    {
        "TenSuKien": "\n          Trọn bộ \"Hot\" Khoa học để cung cấp nhóm ngành Khoa học Sức khỏe, Sinh học, Môi trường\n",
        "NoiDungSuKien": "\n          Chương trình tư vấn trực tuyến Your Future - Your Choice ra ngày 12/4 vừa qua đã mang đến những thông tin hữu ích về nhóm ngành Khoa học sức khỏe,\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583841-tron-bo-bi-kip-chinh-phuc-dam-me-cung-nhom-nhanh-khoa-hoc-suc-khoe-sinh-hoc-moi-truong",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/734aa3d5d919d67bf3310cd97df64378.jpg"
    }
]'

100 %
Results Messages
key value type
1 [ ] { "TenSuKien": "\nYour Future - Your Choice số 10: Lên \"chiến lược\" xét tuyển đại học thông minh\n", "NoiDungSuKien": "\nSau last chủ đề tập trung vào các nhóm ngành cụ thể, chương trình Tư vấn tuyển sinh trực tuyến Your Future - Your Choice số 10 sẽ được phát sóng.\n", "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583886-young-future-your-choice-so-10-len-chien-luoc-xet-tuyen-dai-hoc-thong-minh", "AnhSuKien": "https://file1.hutech.edu.vn/file/news/4eab-ff9d0878b85699a3e22651a231.png" }

Query executed successfully.
Ln 346 Col 6 Ch 6 INS
Item(s) Saved

```

Hình 4.31 Câu lệnh SQL đưa dữ liệu từ JSON lên hệ thống 1

```

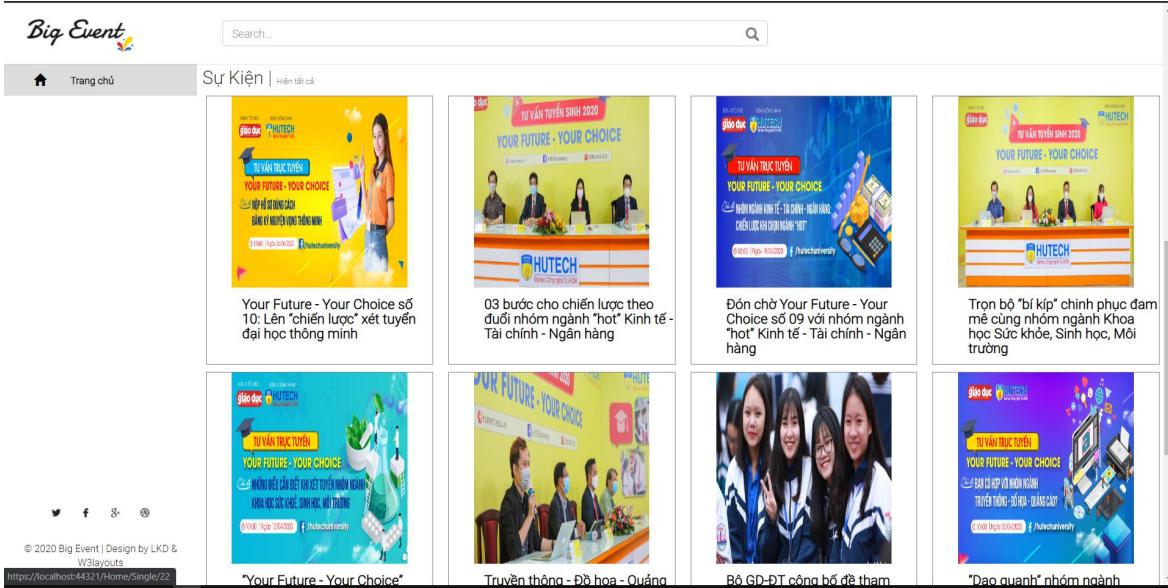
GO
--Nap dữ liệu JSON lên hệ thống
DECLARE @json NVARCHAR(MAX) = N'[
    {
        "TenSuKien": "\nYour Future - Your Choice số 10: Lên \"chiến lược\" xét tuyển đại học thông minh\n",
        "NoiDungSuKien": "\nSau last chủ đề tập trung vào các nhóm ngành cụ thể, chương trình Tư vấn tuyển sinh trực tuyến Your Future - Your Choice số 10 sẽ được phát sóng.\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583886-young-future-your-choice-so-10-len-chien-luoc-xet-tuyen-dai-hoc-thong-minh",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/4eab-ff9d0878b85699a3e22651a231.png"
    },
    {
        "TenSuKien": "\n          03 bước cho chiến lược theo đuổi nhóm ngành \"hot\" Kinh tế - Tài chính - Ngân hàng\n",
        "NoiDungSuKien": "\n          Chương trình tư vấn trực tuyến Your Future - Your Choice số 19/4 vừa qua đã mang đến cho các bạn học sinh yêu thích nhóm ngành Kinh tế - Tài chính\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583882-03-buoc-cho-chien-luoc-theo-duoi-nhom-nhanh-hot-kinh-te-tai-chinh-ngan-hang",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/75c4e8ffa87ca373b95265e7b49a2b7.jpg"
    },
    {
        "TenSuKien": "\n          Bốn chờ Your Future - Your Choice số 09 với nhóm ngành \"hot\" Kinh tế - Tài chính - Ngân hàng\n",
        "NoiDungSuKien": "\n          Với chủ đề Nhóm ngành Kinh tế - Tài chính - Ngân hàng: Chiến lược khi chọn ngành \"hot\", chương trình Tư vấn trực tuyến hứa hẹn sẽ hỗ trợ giải đáp\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583858-don-chon-your-future-your-choice-so-09-voi-nhom-nhanh-hot-kinh-te-tai-chinh-ngan-hang",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/3b04uiddf1fa7faccfesfb03cb3879.png"
    },
    {
        "TenSuKien": "\n          Trọn bộ \"Hot\" Khoa học để cung cấp nhóm ngành Khoa học Sức khỏe, Sinh học, Môi trường\n",
        "NoiDungSuKien": "\n          Chương trình tư vấn trực tuyến Your Future - Your Choice ra ngày 12/4 vừa qua đã mang đến những thông tin hữu ích về nhóm ngành Khoa học sức khỏe,\n",
        "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583841-tron-bo-bi-kip-chinh-phuc-dam-me-cung-nhom-nhanh-khoa-hoc-suc-khoe-sinh-hoc-moi-truong",
        "AnhSuKien": "https://file1.hutech.edu.vn/file/news/734aa3d5d919d67bf3310cd97df64378.jpg"
    }
]'

100 %
Results Messages
key value type
1 [ ] { "TenSuKien": "\nYour Future - Your Choice số 10: Lên \"chiến lược\" xét tuyển đại học thông minh\n", "NoiDungSuKien": "\nSau last chủ đề tập trung vào các nhóm ngành cụ thể, chương trình Tư vấn tuyển sinh trực tuyến Your Future - Your Choice số 10 sẽ được phát sóng.\n", "DuongDanSuKien": "https://www.hutech.edu.vn/tuyensinh/tin-tuc/tin-tuyen-sinh/14583886-young-future-your-choice-so-10-len-chien-luoc-xet-tuyen-dai-hoc-thong-minh", "AnhSuKien": "https://file1.hutech.edu.vn/file/news/4eab-ff9d0878b85699a3e22651a231.png" }

Query executed successfully.
Ln 339 Col 1 Ch 1 INS
Ready

```

Hình 4.32 Câu lệnh SQL đưa dữ liệu từ JSON lên hệ thống 2

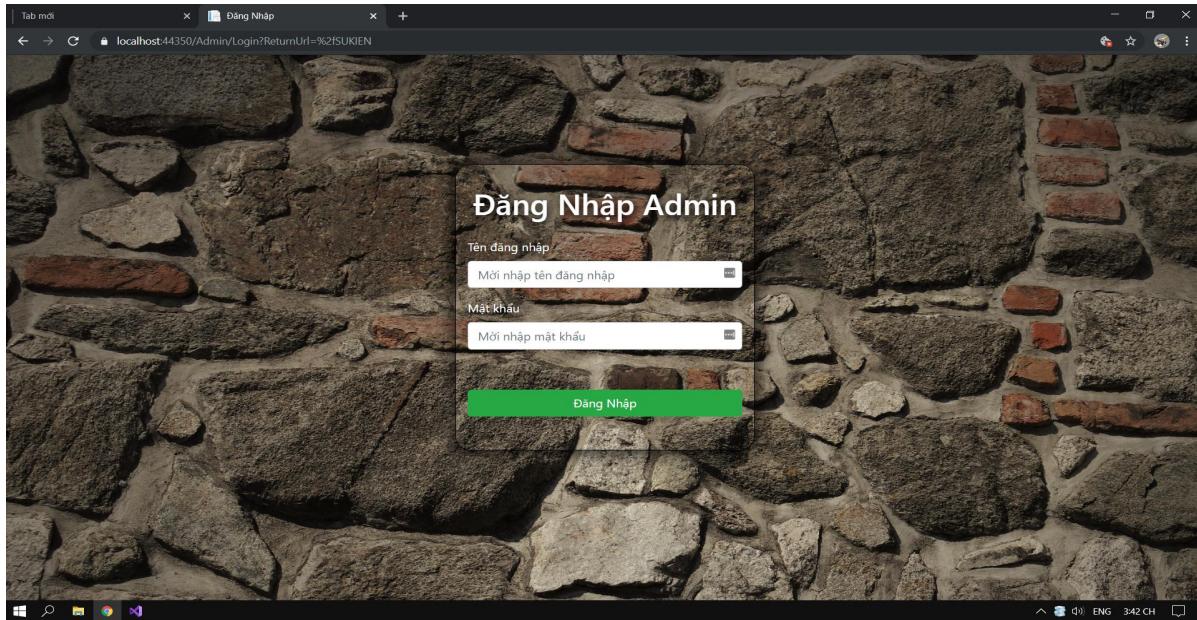


Hình 4.33 Dữ liệu từ JSON đã lên hệ thống 1

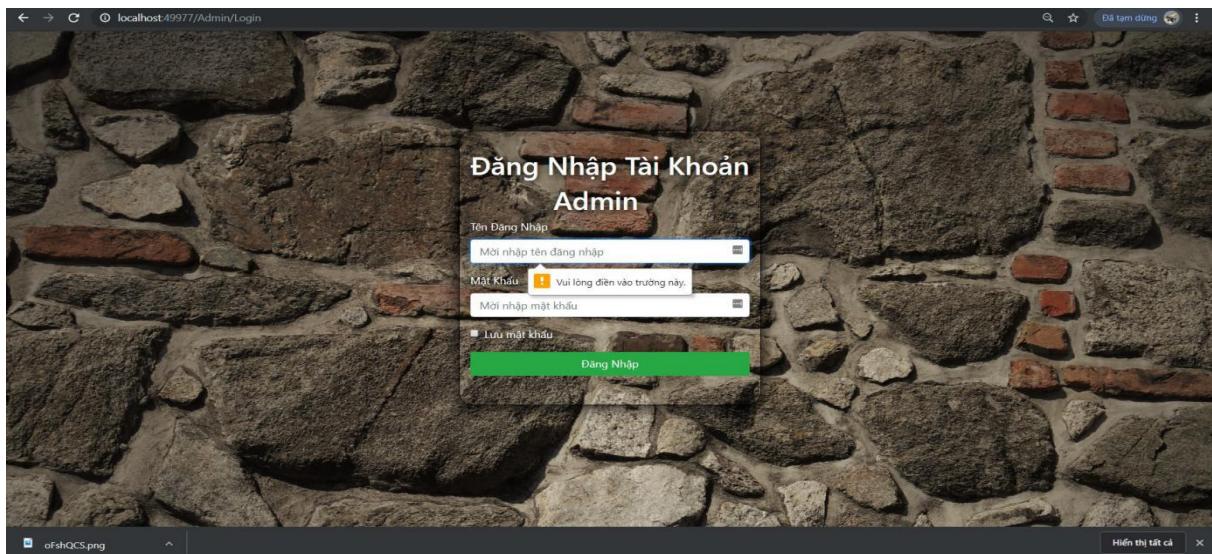
Đánh giá kết quả: Hiện tại về phía bạn làm về Crawl Data vẫn chưa lấy được nhiều thuộc tính có liên quan nên dữ liệu rất là ít, bên SQL Server 2018 không hỗ trợ định dạng UTF8 khi nạp file JSON từ ngoài vào nên gây lỗi font chữ khi sử dụng chữ có dấu, nên phải thêm 1 bước thủ công gây nhập nhằng và chưa được tối ưu.

4.3 Chức năng của hệ thống

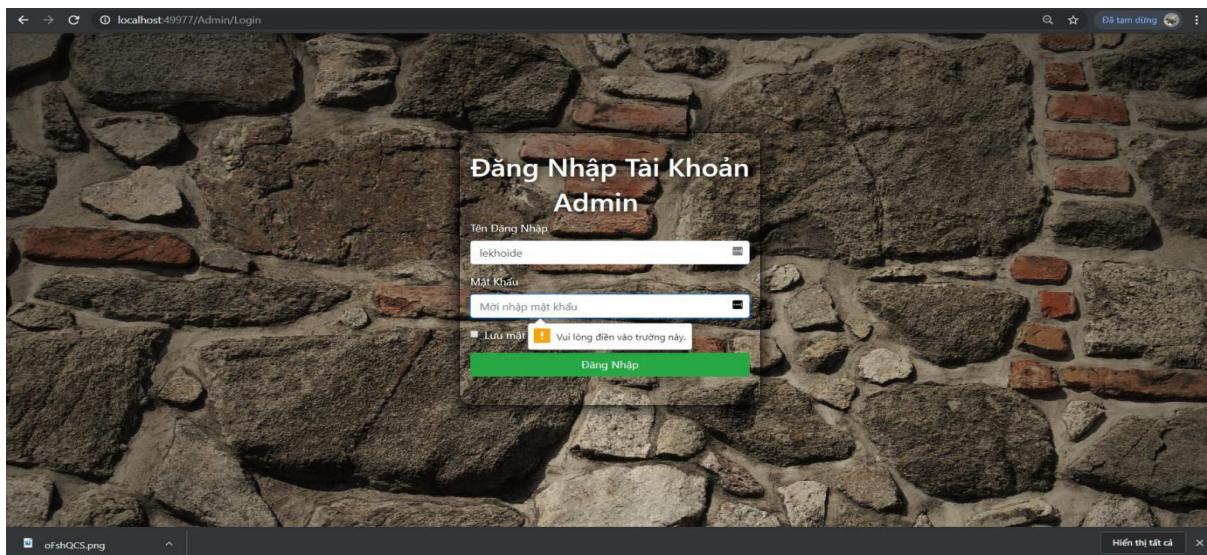
4. 3. 1 Đăng nhập Admin



Hình 4.34 Giao diện đăng nhập Admin 01

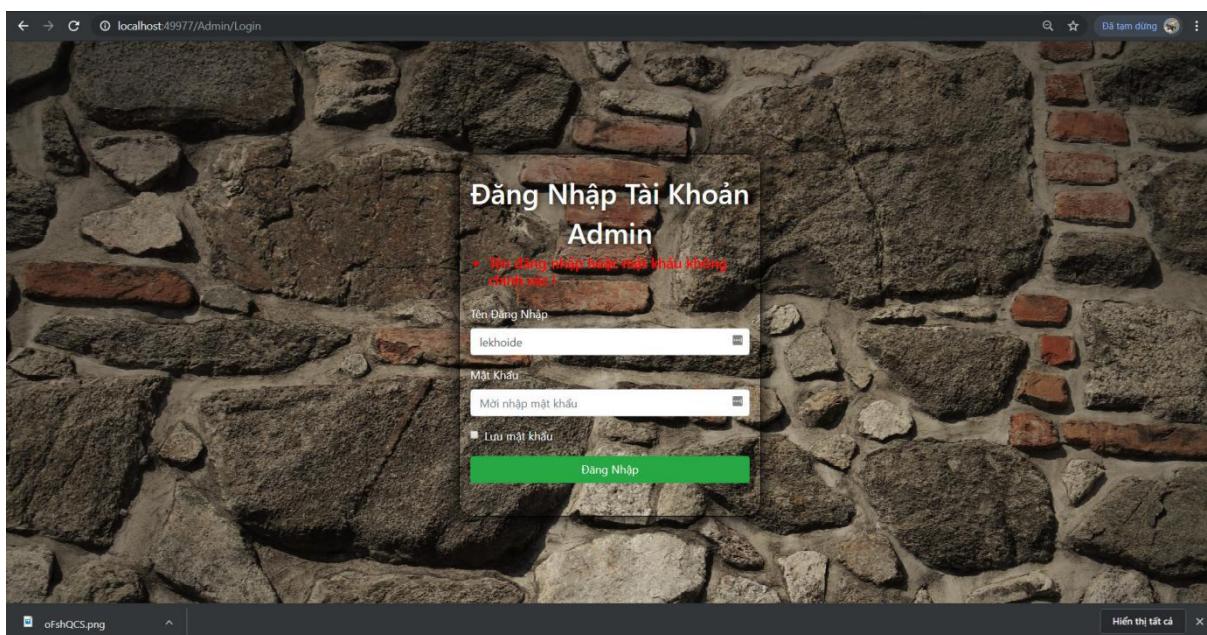


Hình 4.35 Giao diện đăng nhập Admin 02



Hình 4.36 Giao diện đăng nhập Admin 03

-Kết quả trả về khi nhập thông tin không chính xác



Hình 4.37 Giao diện đăng nhập Admin 04

- Kết quả trả về khi đăng nhập thành công

Ảnh sự kiện	Tên sự kiện	Loại sự kiện	Tên bộ phận
	Ngô Hoa	Trong trường	Phòng - Ban Tài chính Phòng - Ban Công tác Sinh viên Phòng - Ban Đảng bộ
	Nắng trưa	Ngoài trường	Khoa Dược Khoa Tiếng Anh Khoa Kiến trúc - Mỹ thuật

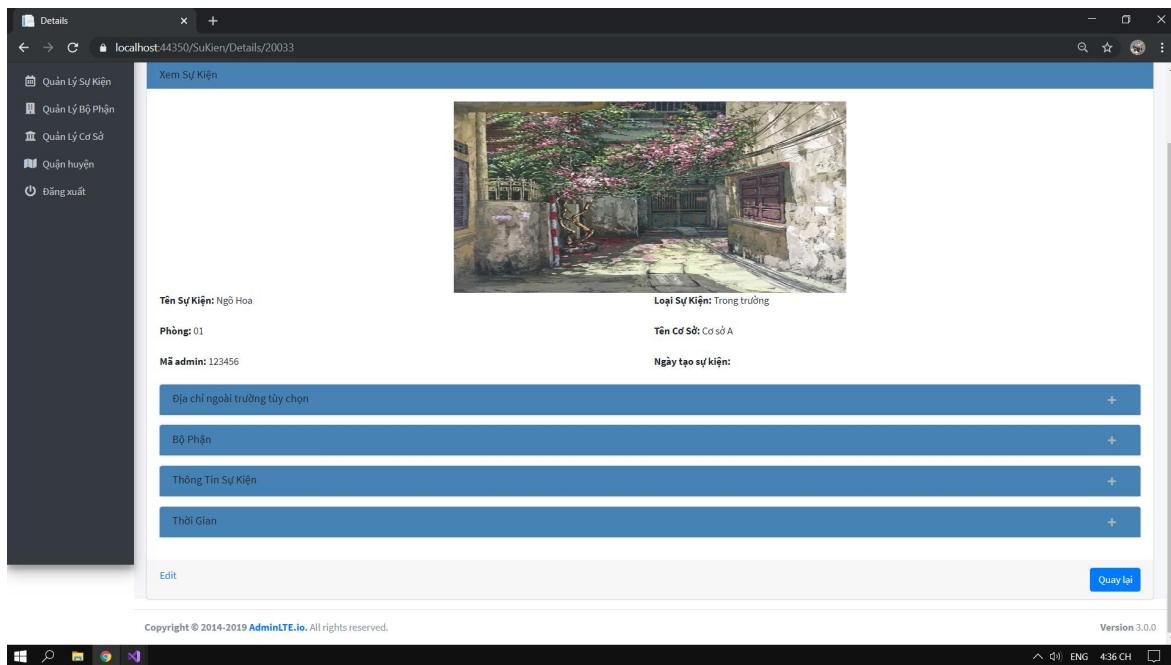
Hình 4.38 Giao diện khi đăng nhập tài khoản thành công

4. 3. 2 CRUD Quản lý sự kiện

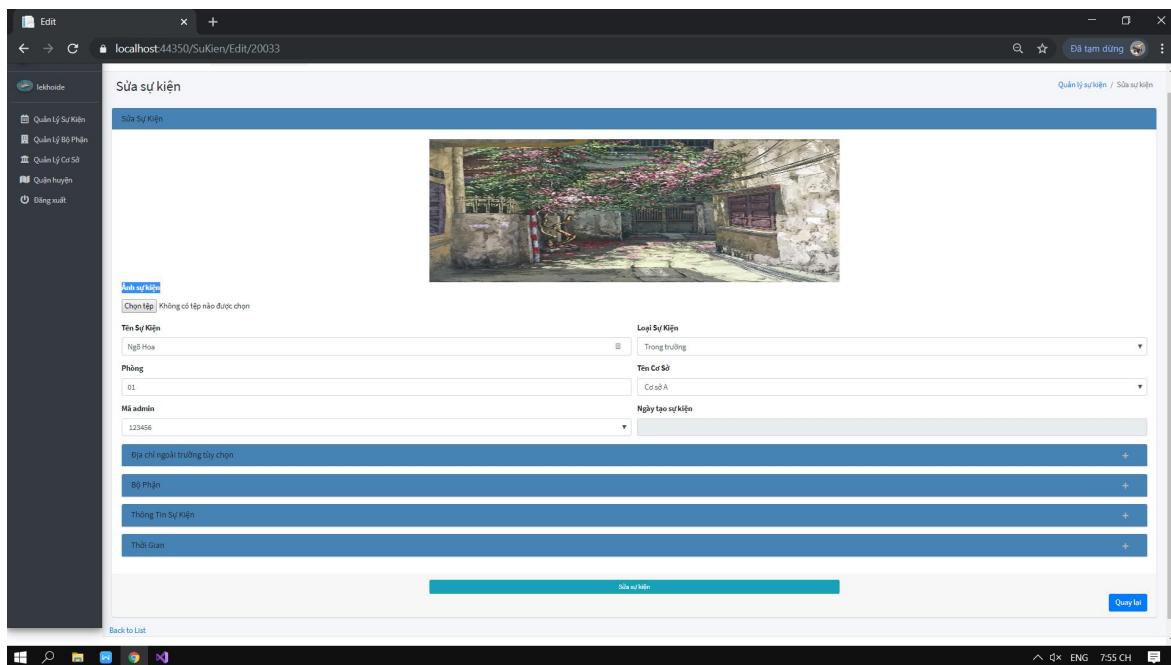
Ảnh sự kiện	Tên sự kiện	Loại sự kiện	
	Thu xưa	Ngoài trường	Xem Sửa Xóa
	Đường hoa	Trong trường	Xem Sửa Xóa
	Đầu xuân	Trong trường	Xem Sửa Xóa

Hình 4.39 Giao diện danh sách sự kiện admin

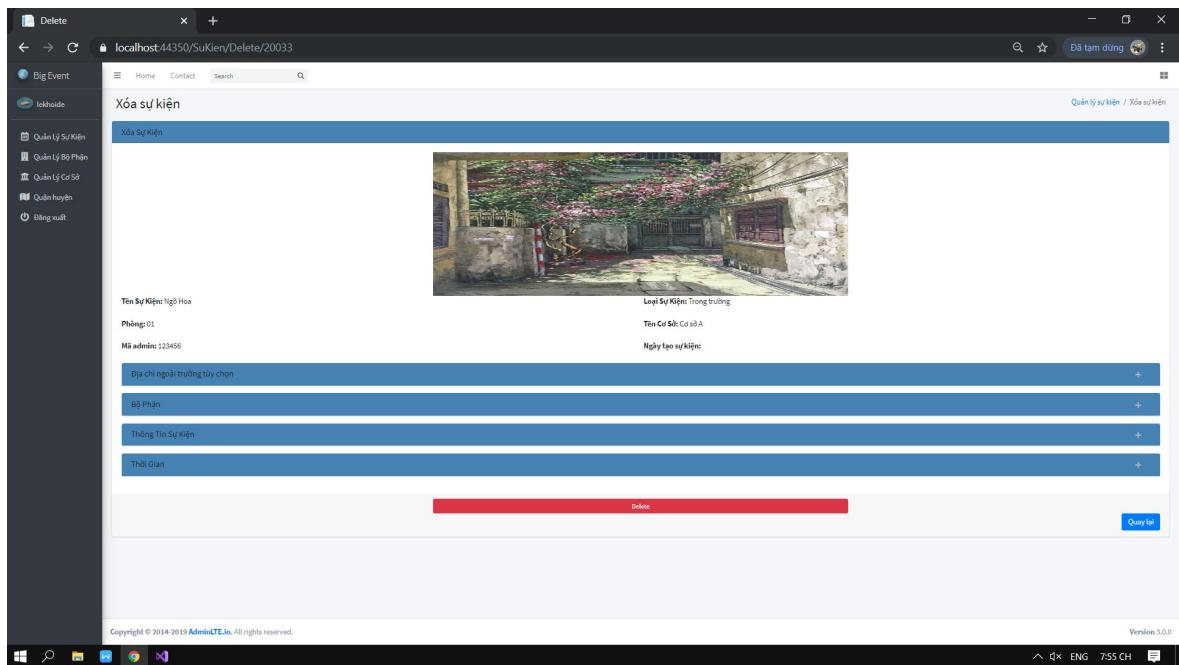
Hình 4.40 Giao diện tạo sự kiện



Hình 4.41 Giao diện xem chi tiết sự kiện admin

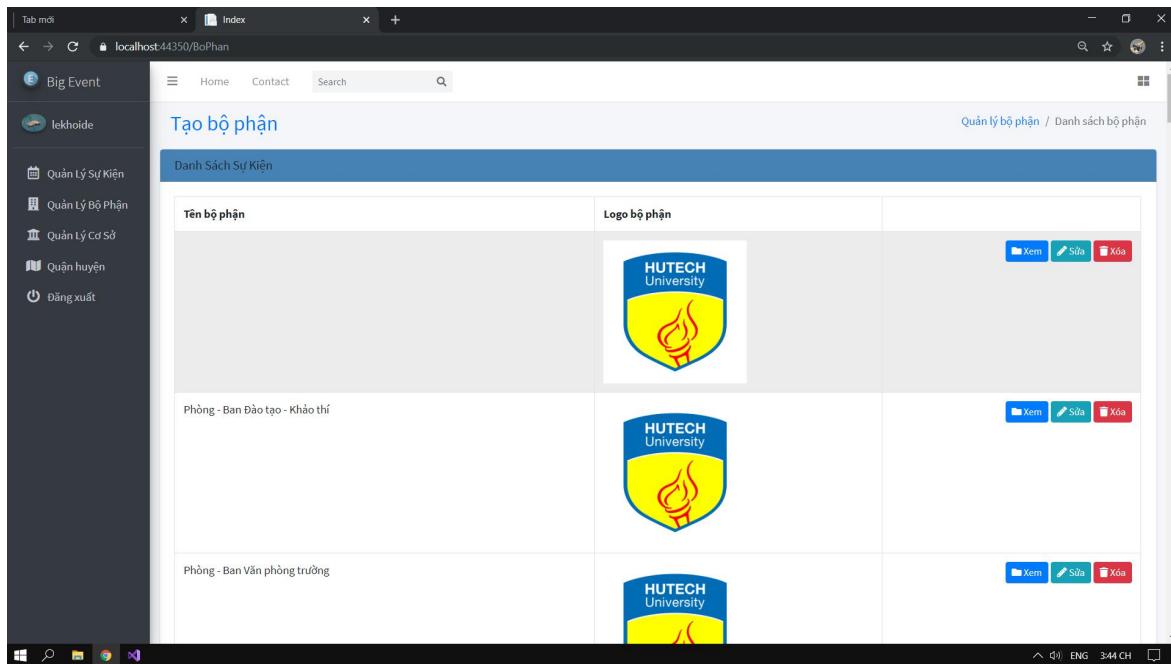


Hình 4.42 Giao diện sửa sự kiện admin



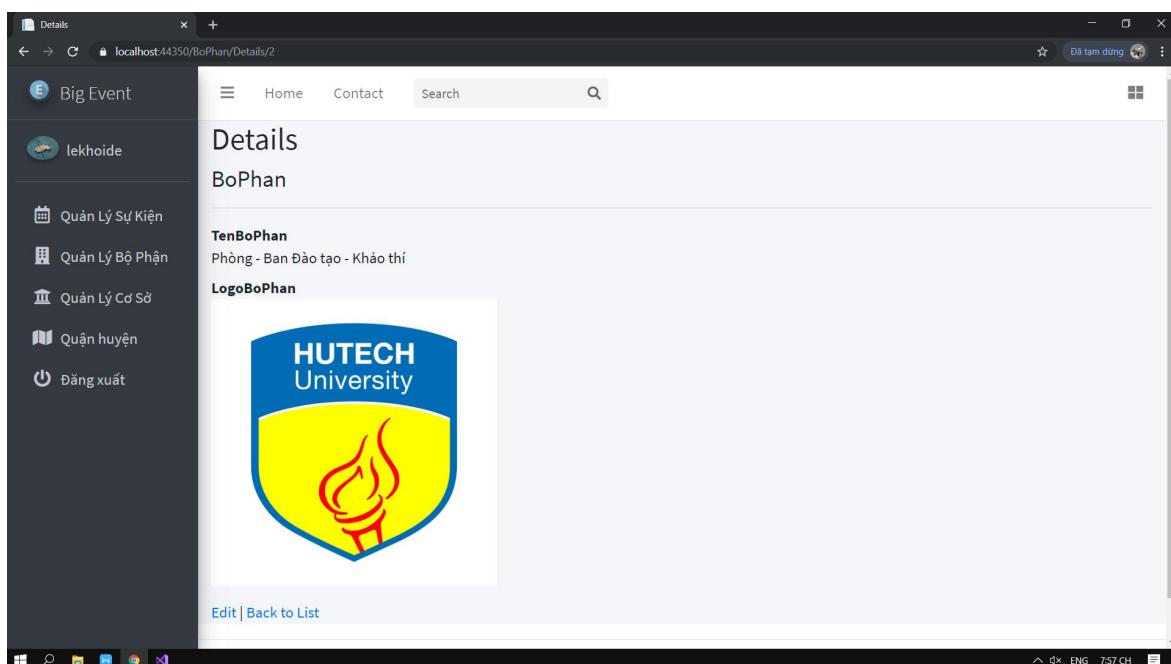
Hình 4.43 Giao diện xóa sự kiện admin

4. 3. 3 CRUD Quản lý bộ phận



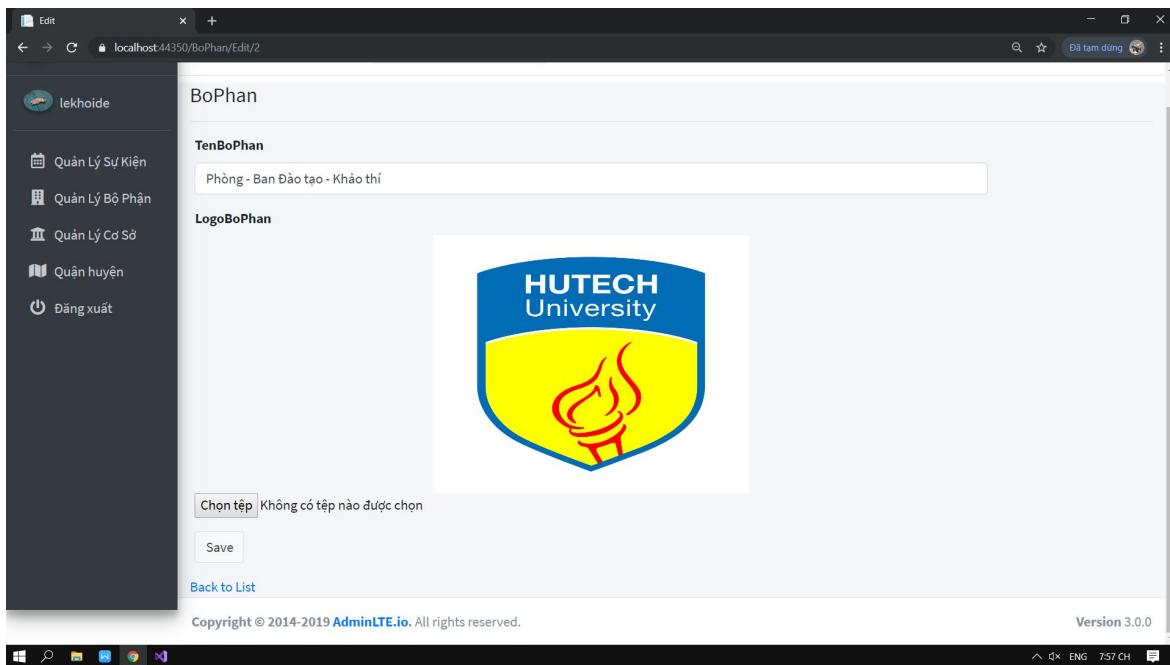
The screenshot shows a web browser window titled 'Index' at 'localhost:44350/BoPhan'. The left sidebar has a dark theme with icons for Big Event, lekhoide, and several menu items: Quản Lý Sự Kiện, Quản Lý Bộ Phận, Quản Lý Cơ Sở, Quận huyện, and Đăng xuất. The main content area is titled 'Tạo bộ phận' (Create Department) and shows a table titled 'Danh Sách Bộ Phận' (List of Departments). The table has columns 'Tên bộ phận' (Department Name) and 'Logo bộ phận' (Department Logo). It lists three departments: 'Phòng - Ban Đào tạo - Khảo thí' with logo 'HUTECH University', 'Phòng - Ban Văn phòng trường' with logo 'HUTECH University', and 'Phòng - Ban Điều hành' with logo 'HUTECH University'. Each row has 'Xem' (View), 'Sửa' (Edit), and 'Xóa' (Delete) buttons.

Hình 4.44 Giao diện danh sách bộ phận admin

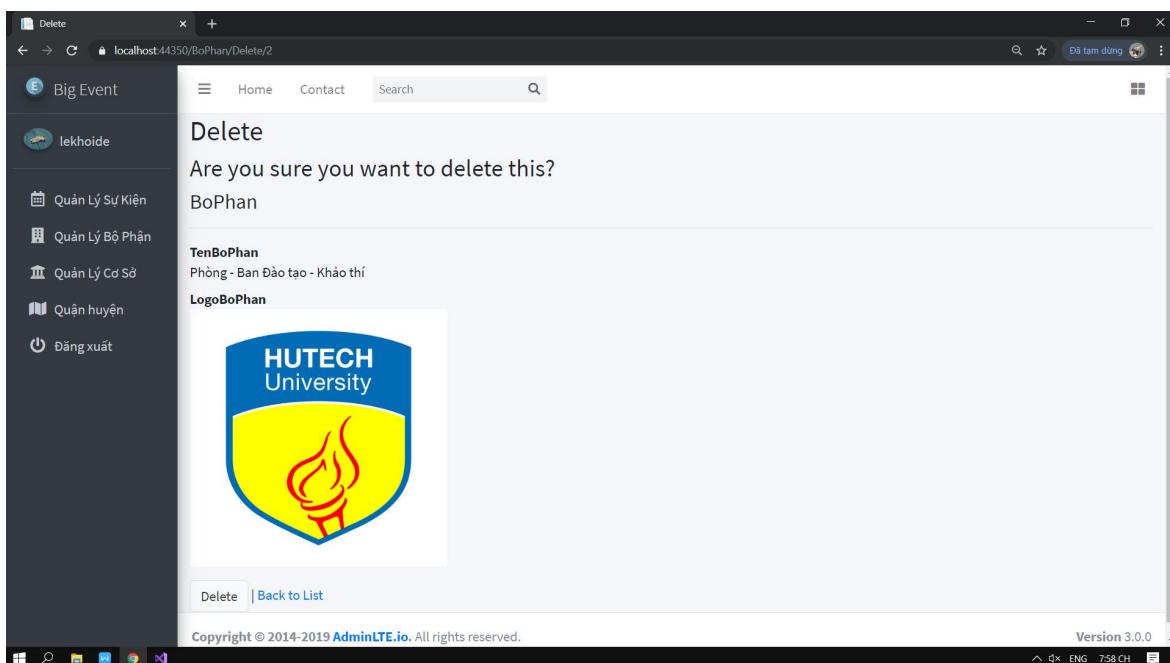


The screenshot shows a web browser window titled 'Details' at 'localhost:44350/BoPhan/Details/2'. The left sidebar is identical to the previous screenshot. The main content area is titled 'Details' and 'BoPhan'. It shows a section 'TenBoPhan' with the value 'Phòng - Ban Đào tạo - Khảo thí' and a section 'LogoBoPhan' showing the 'HUTECH University' logo. At the bottom are 'Edit' and 'Back to List' links.

Hình 4.45 Giao diện chi tiết bộ phận admin



Hình 4.46 Giao diện sửa bộ phận admin



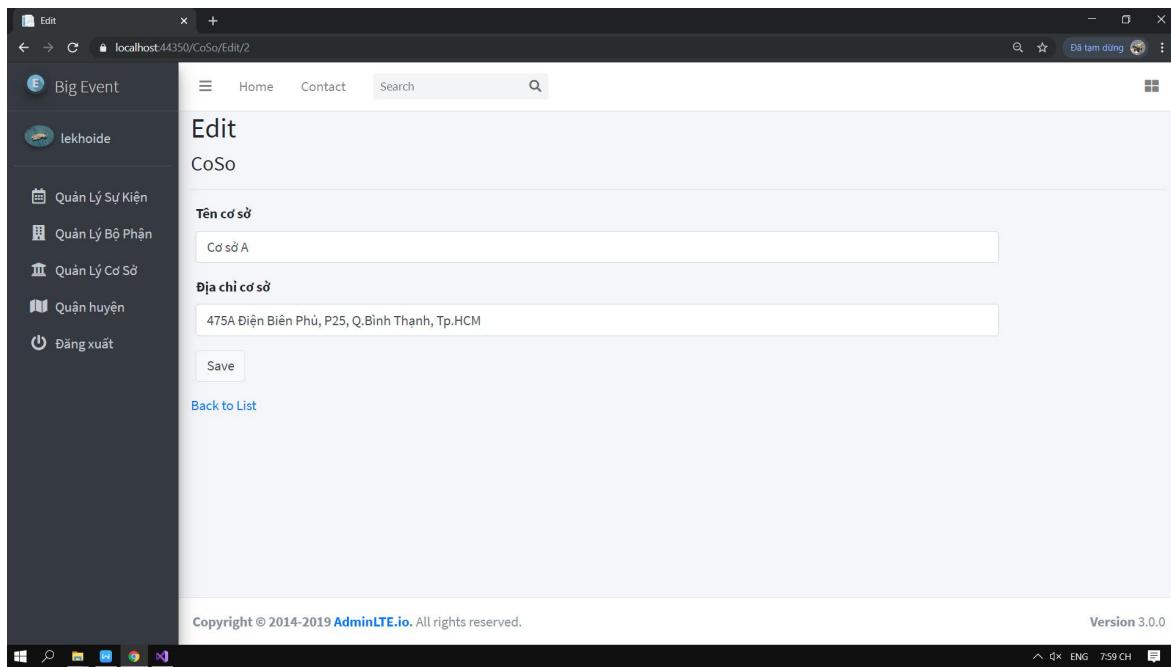
Hình 4.47 Giao diện xóa bộ phận admin

4. 3. 4 CRUD Quản lý cơ sở

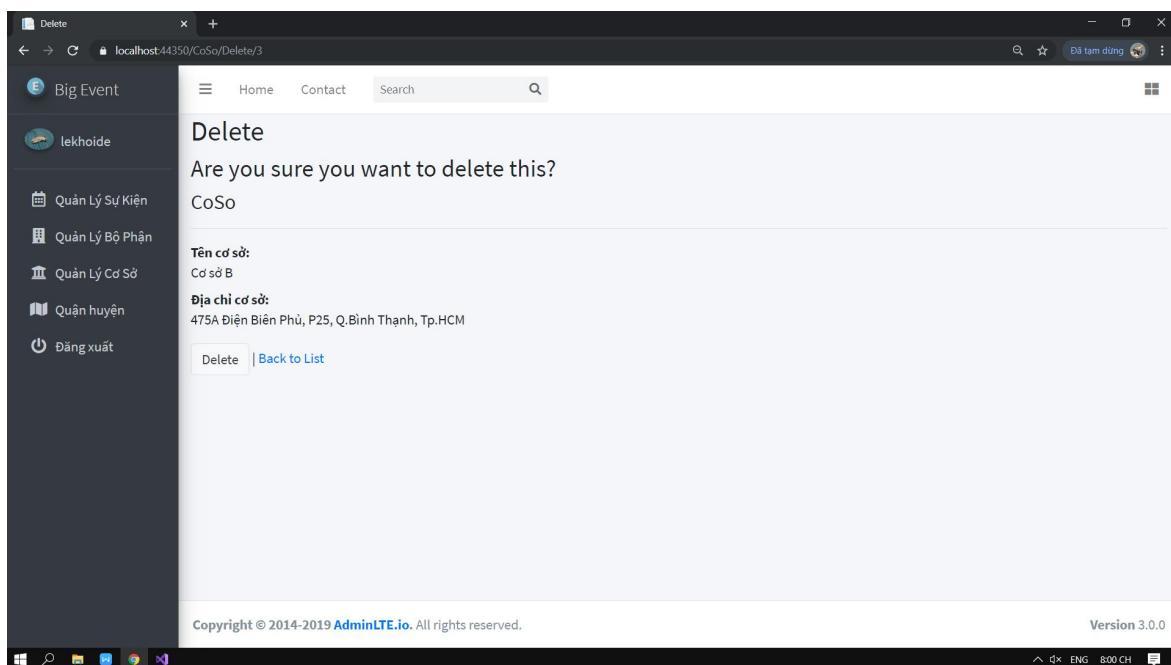
Tên cơ sở	Địa chỉ cơ sở	Xem	Sửa	Xóa
Cơ sở A	475A Điện Biên Phủ, P25, Q.Bình Thạnh, Tp.HCM			
Cơ sở B	475A Điện Biên Phủ, P25, Q.Bình Thạnh, Tp.HCM			
Cơ sở U	31/36 Ưng Văn Khiêm, P.25, Q.Bình Thạnh, Tp.HCM			
Cơ sở E	Lô E1, Phân khu đào tạo E1, Khu Công Nghệ cao TP.HCM, Phường Hiệp Phú, Quận 9, TP.HCM			
Cơ sở R	Viện Công nghệ Cao Hutech, Lô E2B4, đường D1, Phường Long Thạnh Mỹ, khu Công Nghệ Cao, Quận 9, TP.HCM			
Tên cơ sở	Địa chỉ cơ sở			

Hình 4.48 Giao diện danh sách cơ sở admin

Hình 4.49 Giao diện xem chi tiết cơ sở admin



Hình 4.50 Giao diện sửa cơ sở admin



Hình 4.51 Giao diện xóa cơ sở admin

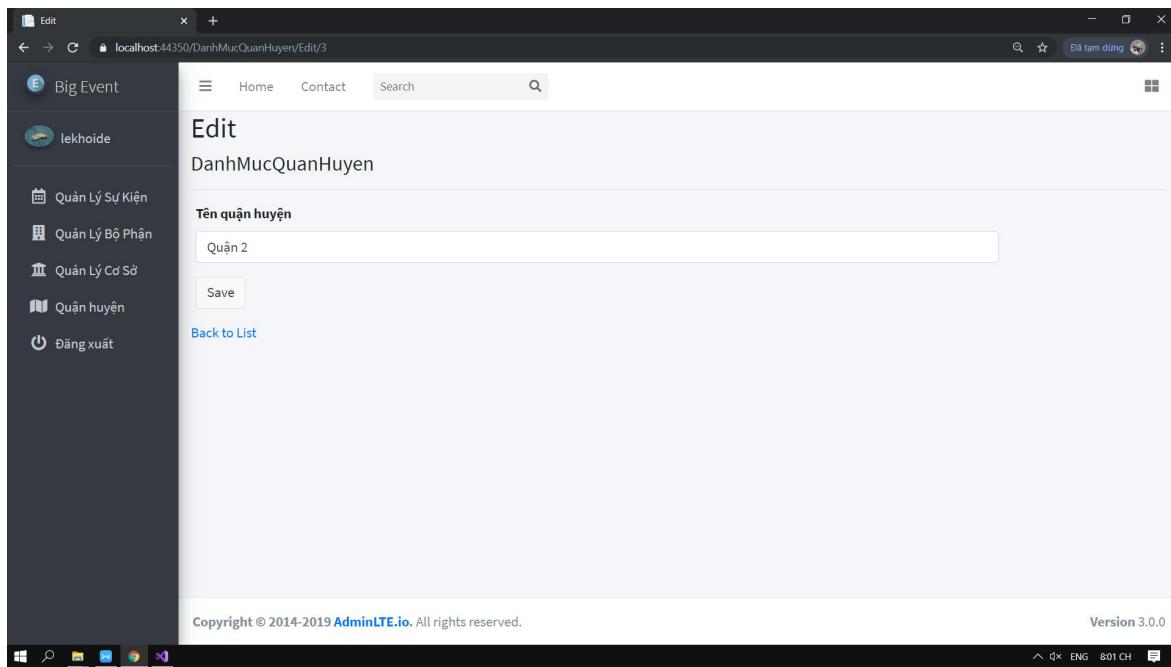
4. 3. 5 CRUD Quản lý danh mục quận huyện

The screenshot shows a web application interface with a dark sidebar menu on the left. The sidebar includes links for 'Index', 'Big Event', 'lekhoidc', 'Quản Lý Sự Kiện', 'Quản Lý Bộ Phận', 'Quản Lý Cơ Sở', 'Quận huyện', and 'Đăng xuất'. The main content area has a header 'Tạo quận huyện' and a breadcrumb 'Quản lý quận huyện / Danh sách quận huyện'. Below this is a table titled 'Danh Sách Quận Huyện' with columns for 'Tên quận huyện' and actions 'Xem', 'Sửa', and 'Xóa'. The table lists seven entries: Quận 1, Quận 2, Quận 3, Quận 4, Quận 5, Quận 6, and Quận 7.

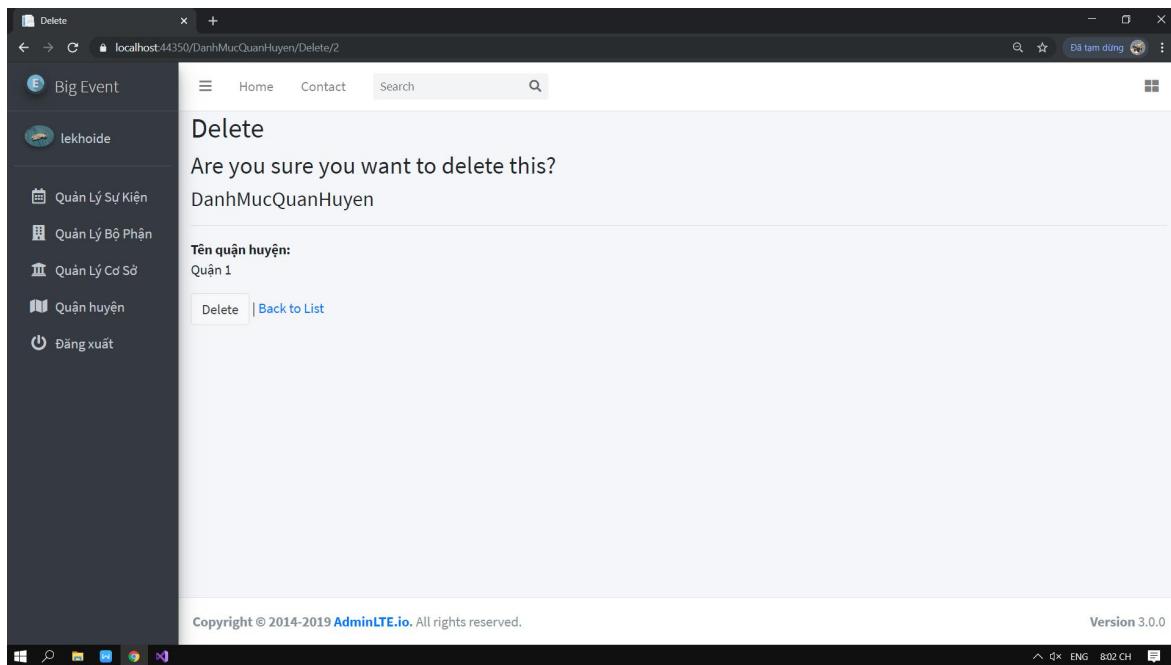
Hình 4.52 Giao diện danh mục quận huyện admin

The screenshot shows the same web application interface as the previous one. The main content area displays a 'Details' section for 'DanhMucQuanHuyen'. It shows the 'Tên quận huyện:' field with the value 'Quận 1'. Below this are 'Edit' and 'Back to List' links. At the bottom of the page, there is a copyright notice 'Copyright © 2014-2019 AdminLTE.io. All rights reserved.' and a version note 'Version 3.0.0'.

Hình 4.53 Giao diện chi tiết quận huyện admin



Hình 4.54 Giao diện sửa quận huyện admin



Hình 4.55 Giao diện xóa quận huyện admin

CHƯƠNG 5 KẾT LUẬN

4. 1 Kết luận chung

Với mục tiêu đề ra ban đầu là thiết kế và xây dựng Website “Công thông tin sự kiện tại Đại Học Công Nghệ - Hutech” trên nền tảng công nghệ ASP.NET MVC hướng dẫn bởi giảng viên Lê Thị Ngọc Thơ trong vòng 10 tuần. Thì nhóm thấy cần phải phát triển đề tài hơn nữa vì còn nhiều bài toán chưa giải quyết và nhóm cần trau dồi kiến thức hơn nữa. Và mong rằng ý tưởng thô sơ mà chúng em thể hiện qua đề tài này sẽ được cô và các thầy chú ý. Nhóm em xin chân thành cảm ơn.

4. 2 Khẳng định kết quả đạt được

4. 2. 1 Kết quả đạt được

- ✧ Xây dựng được một trang Webiste nói chung và hoàn thành đề tài nói riêng, kết quả đạt được không chỉ là một công việc xây dựng một trang Website mà nó là cả một quá trình cố gắng không biết mệt trong vòng 10 tuần của nhóm.
- ✧ Trước một vấn đề về nhu cầu thông tin của sinh viên và sự phát triển mạnh mẽ của công nghệ thông tin thì việc tạo ra một trang Website tin tức riêng giành cho các trường đại học là điều cần thiết và mang lại hiệu quả thực tế cao. Qua việc xây dựng là thực hiện đồ án chuyên ngành cũng giúp chúng em học hỏi thêm được nhiều điều về công việc của lập trình viên và nắm bắt được làm một Website là như thế nào từ đó giúp chúng em có được tiền đề để phát triển thêm nhiều Website khác.
- ✧ Thực hiện được công nghệ Crawl Data từ một Website và đẩy dữ liệu vào được lên hệ thống qua công cụ SQL

4.2.2 Đóng góp

4.2.2.1 Nguyễn Bá Tâm

Em đóng góp được gì vào đồ án môn học này ?

- ◆ Tìm hiểu công nghệ Crawl data
- ◆ Cào dữ liệu về và lưu dưới dạng json

Em học được gì sau khi đồ án chuyên ngành kết thúc ?

- Biết cách phân tích trang web cơ bản
- Học quy trình crawl dữ liệu

Với vai trò trong nhóm em đánh giá mình như thế nào ?

- ◆ Hoàn thành nhiệm vụ ban đầu đề ra

4.2.2.2 Lê Khôi Đê

Em đóng góp được gì vào đồ án môn học này ?

- ❖ Viết đặc tả, thiết kế cơ sở dữ liệu trên SQL Server
- ❖ Vẽ mô hình UseCase Diagram, ClassDiagram, ERD Diagram
- ❖ Code chức năng CRUD Quản lý sự kiện, CRUD Quản lý cơ sở, CRUD Quản lý bộ phận, CRUD danh mục quận huyện
- ❖ Thiết kế giao diện admin : CRUD Quản lý sự kiện, CRUD Quản lý cơ sở, CRUD Quản lý bộ phận, CRUD danh mục quận huyện..
- ❖ Thiết kế giao diện người dùng: trang chủ và xem chi tiết sự kiện
- ❖ Gán giao diện admin (CRUD Quản lý sự kiện, CRUD Quản lý cơ sở, CRUD Quản lý bộ phận, CRUD danh mục quận huyện), người dùng (Trang chủ, Xem chi tiết sự kiện) lên hệ thống
- ❖ Gán giao diện người dùng: trang chủ và xem chi tiết sự kiện lên hệ thống
- ❖ Phân tích Website cần cào ở tuần 8 để có tư liệu viết code bản Demo tuần 8 cho nhóm
- ❖ Nạp file JSON lên hệ thống bằng SQL
- ❖ Viết và chỉnh sửa báo cáo cho nhóm

Em học được gì sau khi đồ án chuyên ngành kết thúc ?

- ❖ Áp dụng được vẽ mô hình trên CaseStudio và xuất code qua SQL Server
- ❖ Áp dụng được EntityDatabase First

- ✧ Kiểm thử quy trình dữ liệu có giống đặc tả trên SQL Server trước khi đưa lên hệ thống
- ✧ Sử dụng JavaScript trên giao diện
- ✧ Học được lý thuyết về công nghệ Crawl Data và quy trình Crawl Data
- ✧ Học được cách viết một dự án Crawl Data cào từ một Website và xuất dữ liệu cào được qua file JSON
- ✧ Học được cách nạp file JSON lên SQL đầy lên hệ thống.

Với vai trò trong nhóm em đánh giá mình như thế nào ?.

- ✧ Vai trò Designer : hoàn thành tốt vì tự tay thiết kế và chỉnh sửa giao diện người dùng theo đúng yêu cầu đặt ra.
- ✧ Vai trò Development : hoàn thành tốt vì làm đúng tiến độ.

4.3. Đánh giá phần mềm

4. 2. 1 Ưu điểm

Có một kênh thông tin dễ dàng giúp các bạn sinh viên tìm kiếm các sự kiện của trường.

Nhiều thông tin sự kiện sẽ được dễ dàng đưa đến nhiều sinh viên hơn thông qua nhiều kênh nhỏ lẻ khác nhau.

4. 2. 2 Nhược điểm

Chức năng tìm kiếm còn là một vấn đề khó khăn.

Giao diện người dùng còn bị lỗi nếu sử dụng nhiều kích thước hiển thị trình duyệt khác nhau.

Dữ liệu cào về chưa được nhiều nên đề tài còn chưa thực sự hoàn chỉnh

4.4. Hướng phát triển và mở rộng đề tài

Đề tài “Cổng thông tin sự kiện tại Đại Học Công Nghệ - Hutech” cần được mở rộng và xem xét trên nhiều khía cạnh hơn để phần mềm được hoàn thiện, giúp cho việc quản lý sự kiện dễ dàng hơn, giảm bớt sự tìm kiếm khi chia sẻ quá nhiều kênh khác nhau... Trong đề tài này, chúng em chỉ mới phân tích và xây dựng phần mềm đơn giản, cần phải phát triển và làm rõ thêm:

- ❖ Tích hợp trên điện thoại di động cho việc thuận tiện theo dõi và cập nhật sự kiện trường
- ❖ Địa chỉ ngoài trường tùy chọn có thể tích hợp được Google Maps
- ❖ Chức năng check in tại sự kiện bằng mã QRcode
- ❖ Cào được thêm những dữ liệu khác.
- ❖ Phân lớp dữ liệu.

TÀI LIỆU THAM KHẢO

- [1] <https://freetuts.net/huong-dan-crawl-du-lieu-trang-chonbaihat-bang-ajax-+-php-1666.html>
- [2] <http://viet.jnlp.org/tai-nguyen-ngon-ngu-tieng-viet/ke-hoach-xay-dung-tu-dong-corpus-tu-nguon-web/crawler>
- [3] <https://topdev.vn/blog/crawling-du-lieu-tu-website-tim-hieu-ve-scrapingweb/>
- [4] <https://lptech.asia/dich-vu/dich-vu-lap-trinh-cong-cu-crawl-du-lieu-website-tu-dong>
- [5] <https://toidicodedao.com/2015/07/28/tutorial-trich-xuat-thong-tin-tu-website-voi-html-agility-pack/>
- [6] <https://codetheworld.io/nodejs-dung-selenium-cheerio-de-crawl-du-lieu.html>
- [7] <https://toidicodedao.com/2015/07/28/tutorial-trich-xuat-thong-tin-tu-website-voi-html-agility-pack/>