1. **Mục lục (Table of Contents)**: Liệt kê các phần chính trong báo cáo.
2. **Tóm tắt (Abstract):**

Đồ án về việc crawler dữ liệu về các hội nghị về lĩnh vực công nghệ thông tin trên thế giới là một dự án có thể được thực hiện bằng việc sử dụng các kỹ thuật và công cụ để tự động thu thập thông tin từ các trang web và cơ sở dữ liệu liên quan đến các hội nghị quốc tế.

1. **Giới thiệu (Introduction)**:
   * Mục tiêu chính của dự án này là xây dựng một hệ thống crawler dữ liệu để thu thập thông tin về các hội nghị uy tín trên toàn cầu. Các thông tin này có thể bao gồm tên hội nghị, các mốc deadline, ngày diễn ra, địa điểm, chủ đề và các thông mở rộng khác như: danh sách diễn giả, các bài báo đã được chấp nhận, và thông tin liên quan đến đăng ký tham dự,...
   * Đặt vấn đề : Trong thời đại số hóa, việc tìm kiếm thông tin về các hội thảo và sự kiện trong lĩnh vực công nghệ thông tin (IT) ngày càng trở nên quan trọng. Tuy nhiên, việc tìm kiếm thông tin này có thể gặp khó khăn do sự phân tán của các nguồn thông tin và sự đa dạng của các hội thảo.
   * Lý do thực hiện đồ án: Xây dựng một trang web **IT Conferences Crawler** Thuận tiện cho người dùng để tiết kiệm thời gian và tối ưu hóa việc tìm kiếm thông tin về các hội thảo, cung cấp thông tin chính xác và tin cậy về các hội thảo trong lĩnh vực công nghệ thông tin giúp người dùng nắm bắt cơ hội học hỏi và giao lưu.
   * Khó khăn và thách thức:

* Một số trang web có chống crawl hoặc bị giới hạn về tốc độ truy cập.
* Dữ liệu trên các trang web có thể thay đổi thường xuyên, đòi hỏi cập nhật định kỳ.
* Dữ liệu có thể không được trình bày trong định dạng dễ phân tích.

1. **Phân tích yêu cầu (Requirements Analysis)**:

* **Mục tiêu của đồ án:** Xây dự một trang web **IT Conferences Crawler** để giúp người dùng tìm kiếm, theo dõi và đăng ký tham gia các hội thảo trong lĩnh vực công nghệ thông tin
* **Yêu cầu chức năng:**

Tìm kiếm hội thảo: người dùng có thể tìm kiếm hội thảo theo tên hội thảo, diễn giả, thời gian, chủ đề ,...v…v

Xem danh sách hội thảo: hiển thị danh sách hội thảo dựa trên tìm kiếm hoặc danh mục hoặc danh mục

Xem thông tin chi tiết hội thảo: Hiển thị thông tin chi tiết về hội thảo. Bao gồm: tên hội nghị, các mốc deadline, ngày diễn ra, địa điểm, chủ đề và các thông tin mở rộng khác như: danh sách diễn giả, các bài báo đã được chấp nhận, và thông tin liên quan đến đăng ký tham dự,...

Xử lý dữ liệu: Sau khi thu thập dữ liệu, cần xử lý và lọc thông tin để loại bỏ dữ liệu không cần thiết và đảm bảo độ chính xác của dữ liệu.

Lưu trữ dữ liệu: Dữ liệu thu thập được cần được lưu trữ trong một cơ sở dữ liệu hoặc tệp để có thể sử dụng sau này.

* **Yêu cầu phi chức năng:** Dữ liệu phải được cập nhật thường xuyên để đảm bảo tính chính xác, trình bày theo định dạng dễ phân tích
* **Yêu cầu kĩ thuật:**
* Ngôn ngữ lập trình: Dùng ngôn ngữ lập trình để xây dựng website
* Cơ sở dữ liệu: Lưu trữ thông tin về hội thảo
* **Yêu cầu giao diện:** xây dựng một giao diện người dùng để tương tác với dữ liệu thu thập được, thân thiện và dễ dùng đảm bảo cho việc người dùng dễ dàng tìm kiếm thông tin hội thảo
* **Yêu cầu kiểm thử:** Đảm bảo tích hợp các chức năng hoạt động ổn định
* **Yêu cầu quản lý đồ án:** Đảm bảo được triển khai đúng hạn
* Danh sách các trang web/ nguồn dữ liệu thu thập thông tin:

<https://techcrunch.com/>

<https://research.com/upcoming-conferences/computer-science>

<https://icrcet.org/>

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

[**https://www.icaset.in/**](https://www.icaset.in/)

**By the hand:**

<https://www.blackhat.com/asia-24/>

<https://www.blackhat.com/us-24/>

<https://techcrunch.com/events/tc-disrupt-2024/>

[**https://hotsauce.hotjar.com/**](https://hotsauce.hotjar.com/)

[**https://philadelphiapact.com/**](https://philadelphiapact.com/)

**https://www.venturelab.ca/search?query=innovation+space**

1. **Thiết kế (Design)**:
   * **Mô tả kiến trúc hệ thống.**

Trang web được xây dựng dựa trên kiến trúc client - server, framework NextJs: Người dùng truy cập thông qua giao diện người dùng (client) - máy chủ (server) xử lý yêu cầu và trả về dữ liệu.

Các thành phần chính:

Front-end: Sư dụng công nghệ ReactJs để hiển thị giao diện cho người dùng(framework NextJs)

Back-end: Sử dụng ngôn ngữ lập trình Nodejs để xử lý logic và dữ liệu (framework NextJs)

Cơ sở dữ liệu: Lưu trữ thông tin về hội thảo và các dữ liệu liên quan khác với Google FireBase để sử dụng sau này.

Các dịch vụ bên ngoài: Google FireBase để lưu trữ dữ liệu , Hosting VercelThiết kế giao diện người dùng.

* **Phân tích thiết kế hệ thống:**

*Các tác nhân chính tham gia vào hệ thống:*

* **Người dùng:** Là người sử dụng trang web có nhu cầu tìm kiếm thông tin về các hội thảo, xem danh sách các hội thảo
* **Admin:** là người quản lý hệ thống , có quyền truy cập và thức hiện các chức năng quản lý (trích xuất dữ liệu, xử lý dữ liệu, điều chỉnh thời gian chạy tự động)

*Các UseCase chính tham gia hệ thống:*

* **Người dùng:**

Tìm kiếm hội thảo

Xem danh sách hội thảo

Xem thông tin chi tiết về hội thảo

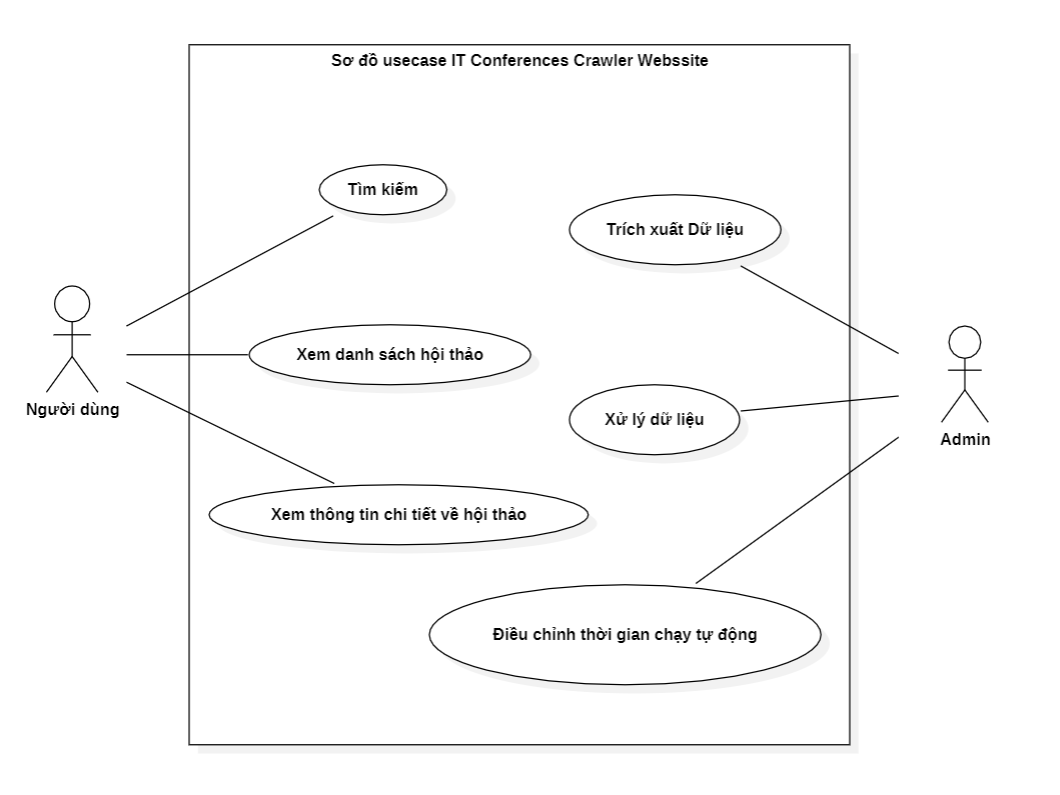
* **Admin:**

Trích xuất dữ liệu

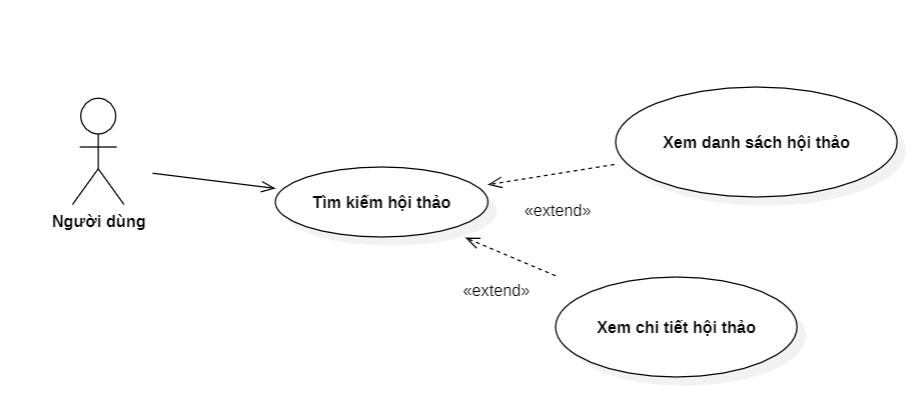
Xử lý dữ liệu

Điều chỉnh thời gian chạy tự động

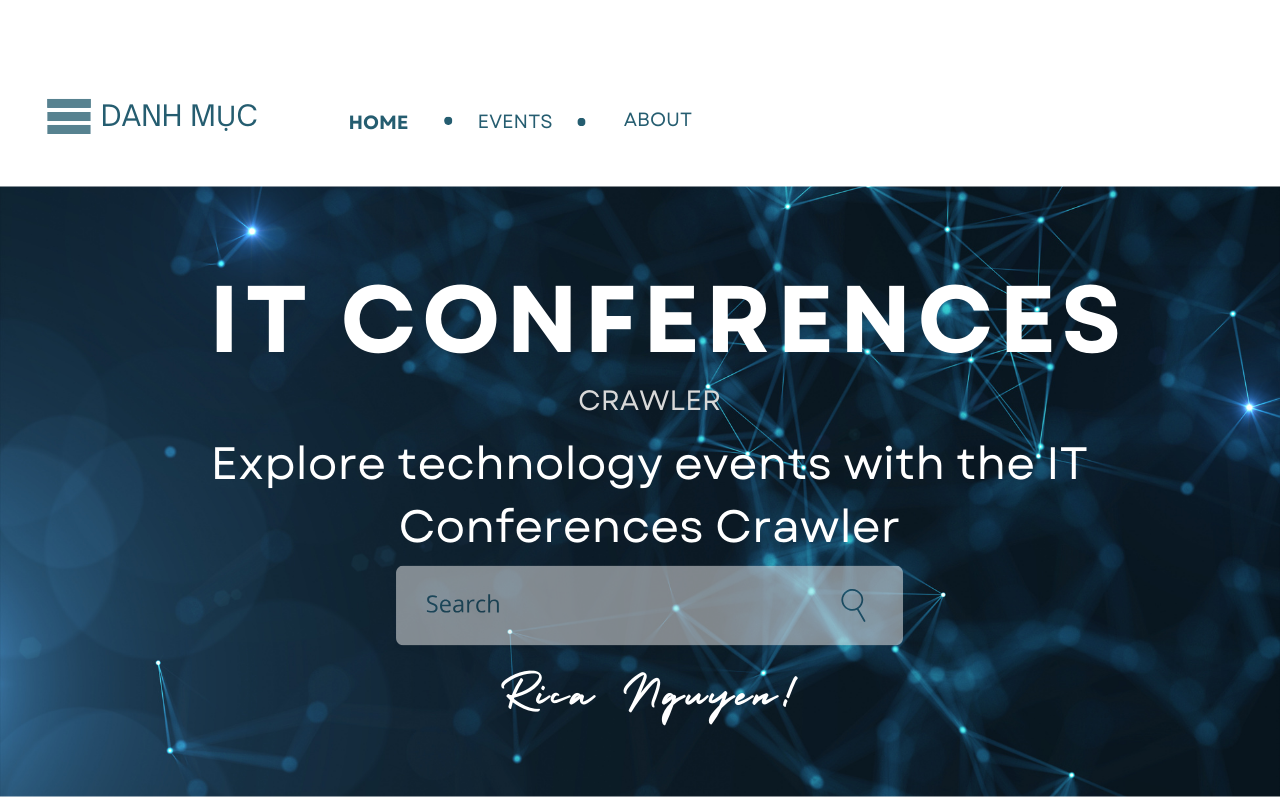
*Lược đồ UseCase tổng quát:*



**Usecase tìm kiếm**

****

**Thiết kế giao diện:**

******

1. **Triển khai (Implementation)**:
   * Mã nguồn và cơ sở dữ liệu.
2. **Kiểm thử (Testing)**:
   * Mô tả quá trình kiểm thử hệ thống.
3. **Kết quả (Results)**:
   * Hiển thị kết quả sau khi triển khai và kiểm thử.
4. **Kết luận (Conclusion)**:
   * Tổng kết đồ án.
   * Đánh giá hiệu suất và hạn chế của hệ thống.
5. **Tài liệu tham khảo (References)**:
   * Liệt kê các tài liệu đã tham khảo trong quá trình thực hiện đồ án.