IT CONFERECNCES CRAWLER

Mô hình thác nước

– Phân tích yêu cầu

– Thiết kế phần mềm

– Cài đặt phần mềm

– Kiểm nghiệm phần mềm

– Triển khai & bảo trì

PHÂN TÍCH YÊU CẦU

MỤC TIÊU: Xây dựng một hệ thống crawler dữ liệu để thu thập thông tin về các hội nghị uy tính trên toán cầu.

THÔNG TIN: Tên hội nghị, các mốc deadline, ngày diễn ra, địa điểm, chủ đề và các thông tin mở rộng: danh sách diễn giả, các bài báo đã được chấp nhận, và các thông tin liên quan đến đăng ký tham dự,…

LẬP KẾ HOẠCH VÀ THIẾT KẾ HỆ THỐNG

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/)

<div> ngày tháng năm

Hội nghị

<href?> link

**Tìm kiếm hội thảo (Search Conference) - Use Case**

**Mục tiêu:**

* Xây dựng chức năng tìm kiếm hội thảo để người dùng có thể dễ dàng tìm kiếm thông tin về các sự kiện công nghệ thông tin trên toàn thế giới.
* Cung cấp kết quả tìm kiếm chính xác và nhanh chóng.

**Các bước thực hiện:**

1. **Nhập từ khóa tìm kiếm**:
   * Người dùng nhập từ khóa hoặc danh mục liên quan đến hội thảo mà họ quan tâm.
2. **Tìm kiếm trong cơ sở dữ liệu**:
   * Hệ thống tìm kiếm trong cơ sở dữ liệu hội thảo dựa trên từ khóa hoặc danh mục đã nhập.
   * Trả về danh sách các hội thảo phù hợp với yêu cầu tìm kiếm.
3. **Hiển thị kết quả tìm kiếm**:
   * Hiển thị danh sách hội thảo với thông tin chi tiết như tên hội thảo, ngày diễn ra, địa điểm, và chủ đề.
   * Cho phép người dùng xem thông tin chi tiết về từng hội thảo.
4. **Lọc kết quả tìm kiếm (tùy chọn)**:
   * Người dùng có thể lọc kết quả tìm kiếm theo ngày, địa điểm, chủ đề, hoặc loại hội thảo.
5. **Kết thúc use case**:
   * Người dùng có thể chọn hội thảo cụ thể để xem thông tin chi tiết hoặc tiếp tục tìm kiếm.

Hội thoại trên giúp người dùng dễ dàng tìm kiếm thông tin về các sự kiện công nghệ thông tin trên toàn thế giới. 🌟

**Xem danh sách hội thảo (View Conference List) - Use Case**

**Mục tiêu:**

* Xây dựng chức năng hiển thị danh sách hội thảo để người dùng có thể xem thông tin về các sự kiện công nghệ thông tin.

**Các bước thực hiện:**

1. **Truy cập danh sách hội thảo**:
   * Người dùng truy cập mục “Danh sách hội thảo” trên trang web.
   * Hệ thống hiển thị danh sách hội thảo có sẵn.
2. **Hiển thị thông tin hội thảo**:
   * Hệ thống hiển thị danh sách hội thảo với thông tin cơ bản như tên hội thảo, ngày diễn ra, địa điểm, và chủ đề.
3. **Lọc danh sách hội thảo (tùy chọn)**:
   * Người dùng có thể lọc danh sách hội thảo theo ngày, địa điểm, chủ đề, hoặc loại hội thảo.
4. **Xem thông tin chi tiết về hội thảo (tùy chọn)**:
   * Người dùng có thể chọn hội thảo cụ thể để xem thông tin chi tiết về từng sự kiện.
5. **Kết thúc use case**:
   * Người dùng có thể tiếp tục tìm kiếm hội thảo hoặc xem thông tin chi tiết về hội thảo.

Hội thoại trên giúp người dùng dễ dàng xem thông tin về các sự kiện công nghệ thông tin. 🌟

**Xem thông tin chi tiết về hội thảo (View Conference Details) - Use Case**

**Mục tiêu:**

* Xây dựng chức năng hiển thị thông tin chi tiết về một hội thảo cụ thể để người dùng có thể tìm hiểu về nó.

**Các bước thực hiện:**

1. **Chọn hội thảo cụ thể**:
   * Người dùng chọn một hội thảo từ danh sách đã hiển thị hoặc tìm kiếm.
2. **Hiển thị thông tin chi tiết**:
   * Hệ thống hiển thị thông tin chi tiết về hội thảo, bao gồm:
     + Tên hội thảo.
     + Ngày diễn ra.
     + Địa điểm tổ chức.
     + Chủ đề và nội dung hội thảo.
     + Danh sách diễn giả và chuyên gia tham gia.
3. **Kết thúc use case**:
   * Người dùng có thể tiếp tục tìm kiếm hội thảo khác hoặc quay lại danh sách hội thảo.

Hội thoại trên giúp người dùng dễ dàng tìm hiểu thông tin chi tiết về các sự kiện công nghệ thông tin. 🌟

**Usecase “Admin Trích Xuất Dữ Liệu Cho Đồ Án IT Conferences Crawler”**

Trong đồ án **IT Conferences Crawler**, usecase “Admin Trích Xuất Dữ Liệu” có nhiệm vụ thu thập thông tin từ các trang web về hội nghị công nghệ thông tin. Dưới đây là một số bước cơ bản để triển khai usecase này:

1. **Gửi HTTP Request**: Sử dụng thư viện requests trong Python để gửi HTTP request đến trang web hội nghị. Kết quả trả về sẽ là một trang dạng document.
2. **Trích Xuất Link**: Từ trang đã lấy được, trích xuất ra các đường link dẫn đến chi tiết thông tin của các hội nghị. Các đường link này sẽ dẫn đến các trang chứa thông tin về thời gian, địa điểm, chủ đề, diễn giả, và nội dung của hội nghị.
3. **Thu Thập Dữ Liệu**: Dựa vào DOM (Document Object Model) của trang web, thu thập dữ liệu cần thiết. Các thông tin có thể bao gồm tên hội nghị, ngày diễn ra, địa điểm, danh sách diễn giả, và nội dung chi tiết.
4. **Lưu Trữ Dữ Liệu**: Sau khi trích xuất, lưu trữ dữ liệu vào cơ sở dữ liệu hoặc file để sử dụng cho mục đích khác.
5. **Xây Dựng Recommendation System**: Dựa trên dữ liệu thu thập được, bạn có thể xây dựng một hệ thống gợi ý hội nghị cho người dùng.

**Usecase “Admin Điều Chỉnh Thời Gian Chạy Tự Động Cho Đồ Án IT Conferences Crawler”**

Trong đồ án **IT Conferences Crawler**, usecase “Admin Điều Chỉnh Thời Gian Chạy Tự Động” có nhiệm vụ điều chỉnh thời gian chạy tự động của hệ thống. Dưới đây là một số bước cơ bản để triển khai usecase này:

1. **Sử Dụng Task Scheduler**: Trên hệ điều hành Windows, sử dụng **Task Scheduler** để lên lịch thực hiện các tác vụ tự động. Bạn có thể thiết lập thời gian chạy cho việc trích xuất dữ liệu từ các trang web hội nghị.
2. **Xác Định Thời Gian Chạy**: Trong Task Scheduler, bạn có thể thiết lập thời gian chạy hàng ngày, hàng tuần hoặc theo một lịch trình cụ thể. Điều này giúp hệ thống tự động thực hiện việc trích xuất dữ liệu vào thời điểm mong muốn.
3. **Thiết Lập Thông Báo**: Nếu có bất kỳ lỗi hoặc vấn đề nào xảy ra trong quá trình chạy tự động, hệ thống cần thông báo cho admin để kiểm tra và xử lý.
4. **Kiểm Tra Tiến Độ**: Theo dõi tiến độ thực hiện tự động để đảm bảo rằng việc trích xuất dữ liệu diễn ra đúng theo kế hoạch.

**secase “Admin Xử Lý Dữ Liệu Cho Đồ Án IT Conferences Crawler”**

Trong đồ án **IT Conferences Crawler**, usecase “Admin Xử Lý Dữ Liệu” có nhiệm vụ xử lý thông tin thu thập từ các trang web về hội nghị công nghệ thông tin. Dưới đây là một số bước cơ bản để triển khai usecase này:

1. **Tiền Xử Lý Dữ Liệu**: Loại bỏ dữ liệu không cần thiết, xử lý dữ liệu thiếu hoặc không hợp lệ, và chuẩn hóa dữ liệu để đảm bảo tính nhất quán.
2. **Ko làm Trực Quan Hóa Dữ Liệu**: Sử dụng biểu đồ, đồ thị hoặc các công cụ khác để hiển thị thông tin một cách trực quan. Ví dụ: biểu đồ thể hiện số lượng hội nghị theo thời gian, địa điểm, chủ đề, v.v.
3. **Ko làm Phân Tích Dữ Liệu**: Áp dụng các phương pháp phân tích dữ liệu để tìm ra các mẫu, xu hướng, và thông tin hữu ích. Ví dụ: phân tích tần suất xuất hiện của các từ khóa trong nội dung hội nghị.
4. **Lưu Trữ Dữ Liệu**: Sau khi xử lý, lưu trữ dữ liệu vào cơ sở dữ liệu hoặc file để sử dụng cho mục đích khác.
5. **Tối Ưu Hóa Dữ Liệu**: Tùy theo yêu cầu của dự án, bạn có thể tối ưu hóa dữ liệu để giảm kích thước, tăng tốc độ truy vấn, hoặc cải thiện hiệu suất.