**NHẬP MÔN CÔNG NGHỆ PHẦN MỀM**

**Project Proposal**

**Nhóm 14:**

21120505 – Bùi Thị Thanh Ngân

21120531 – Nguyễn Văn Phúc

21120540 – Trần Tôn Bửu Quang

21120600 – Nguyễn Thị Như Ý

C:\Users\tdqua_000\Dropbox\SS-Slides\DeCuong-CDIO\Template CDIO v4.2\Templates\Hinh anh\LogoTruong.png

**Bộ môn Công nghệ phần mềm**

**Khoa Công nghệ thông tin**

**Đại học Khoa học tự nhiên TP HCM**

**MỤC LỤC**

Contents

[Các nội dung chính 2](#_Toc150644617)

[1 Bảng đánh giá thành viên 3](#_Toc150644618)

[2 Phát biểu bài toán sơ lược 4](#_Toc150644619)

[3 Giải pháp đề xuất 5](#_Toc150644620)

[4 Kế hoạch phát triển 10](#_Toc150644621)

[5 Kế hoạch nhân sự và chi phí 12](#_Toc150644622)

**Project Proposal**

# Các nội dung chính

Mục tiêu tài liệu tập trung vào các chủ đề:

* Tạo ra tài liệu Project Proposal.
* Hoàn chỉnh tài liệu Project Proposal với các nội dung:
  + Phát biểu bài toán sơ lược
  + Giải pháp đề xuất
  + Kế hoạch phát triển
  + Kế hoạch nhân sự & chi phí
* Đọc hiểu tài liệu Project Proposal.

# Bảng đánh giá thành viên

|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ Tên** | **% đóng góp (tối đa 100%)** | **Chữ ký** |
| 21120505 | Bùi Thị Thanh Ngân | 25% |  |
| 21120531 | Nguyễn Văn Phúc | 25% |  |
| 21120540 | Trần Tôn Bửu Quang | 25% |  |
| 21120600 | Nguyễn Thị Như Ý | 25% |  |

# Phát biểu bài toán sơ lược

Ngày nay, thông tin trên Internet ngày càng nhiều và đến từ nhiều nguồn đa dạng khác nhau, do đó dẫn đến việc khó kiểm soát và khó kiểm chứng liệu chúng có chính xác và có độ tin cậy cao hay không . Vậy nên để đáp ứng cho nhu cầu tìm kiếm, học hỏi, nắm bắt sự thay đổi và cập nhật công nghệ, tin tức mới trên toàn cầu hàng ngày cho những ai quan tâm tới lĩnh vực công nghệ thông tin, nhóm chúng tôi quyết định xây dựng một website với chức năng thu thập, xử lý dữ liệu và hiển thị trực quan các thông tin về những hội nghị liên quan đến lĩnh vực công nghệ thông tin trên thế giới mang tên Crawler.

Trang web crawler sẽ có chức năng thu thập dữ liệu tự động từ các trang web và cơ sở dữ liệu liên quan về các hội nghị quốc tế về lĩnh vực công nghệ thông tin trên thế giới. Thông tin của một hội nghị được thu thập sẽ bao gồm tên hội nghị, các mốc thời gian trong hội nghị, ngày diễn ra, địa điểm, chủ đề cùng một số thông tin khác như: danh sách diễn giả, các bài báo khoa học đã được chấp nhận, các thông tin về đăng ký tham dự,...

Để đảm bảo về chất lượng nguồn thông tin cung cấp cho người dùng thì sau khi tìm được các trang web, web crawler sẽ tiến hành xem xét đến các tiêu chí như mức độ tin cậy, độ phổ biến, lượng tuy cập, ngày đăng bài, lượng thông tin trên trang web.

Các thông tin đã được lấy về cần được lưu trữ vào một cơ sở dữ liệu hoặc một hệ thống nào đó thuận tiện cho việc truy xuất và sử dụng sau này. Vì dữ liệu lấy từ một trang web có lượng thông tin vô cùng đa dạng, thế nên người phát triển web crawler cần phải tiến hành lọc và tìm ra các thông tin hữu ích mà người dùng muốn thu thập, loại bỏ đi các thông tin nhiễu hoặc bị trùng lặp. Vậy nên sau khi đã thu thập dữ liệu về dự án, hội nghị nào đó thì nhà phát triển cần tiến hành các bước phân tích chuyên sâu để loại bỏ các thông tin không cần thiết và đảm bảo độ chính xác của dữ liệu.

Để đảm bảo nguồn thông tin luôn là mới nhất, phù hợp với xu thế hiện nay và đáp ứng được nhu cầu cập nhật mới của người dùng thì web crawler sẽ thường xuyên cập nhật, liên tục tìm kiếm các trang web khác nhau để trích xuất các thông tin đăng tải mới nhất.

Để mọi đối tượng đều có thể sử dụng thì trang web cần phải được trình bày, thiết kế một cách rõ ràng, đẹp mắt, thuận tiện thao tác cho dù là người mới lần đầu sử dụng. Các thông tin trình bày một cách chi tiết nhưng vẫn đảm bảo dễ nhìn, dễ tìm kiếm và truy xuất. Bên cạnh đó để đảm bảo trải nghiệm tốt cho người dùng thì tốc độ phản hồi của website cũng cần nhanh chóng và chính xác. Ngoài ra website cũng cần đảm bảo có thể xử lý nhiều tác vụ truy cập cùng một lúc mà không làm tăng độ trễ hay thời gian chờ.

Đối với người dùng, khi truy cập vào trang web thì sẽ thấy được các thông tin tổng quan, thông tin cập nhật mới nhất về các hội nghị mà người đó có thể quan tâm. Khi người dùng chọn xem hội nghị nào đó thì thông tin chi tiết sẽ được hiển thị một cách rõ ràng, hoặc có thể dẫn người dùng tới trang web của hội nghị đó. Ngoài ra trang web còn hỗ trợ chức năng tìm kiếm để tra cứu nhanh hơn(tìm kiếm theo tên hội nghị, theo tên diễn giả, theo lĩnh vực cụ thể, theo thời gian... và nhiều tiêu chí khác).

# Giải pháp đề xuất

#### . Phần mềm

#### Danh sách các chức năng phần mềm

|  |  |  |
| --- | --- | --- |
| **Nhu cầu** | **Yêu cầu** | **Chi tiết** |
| Người sử dụng cần tìm kiếm thông tin về các hội nghị với từ khóa cho trước | Chức năng tra cứu, tìm kiếm | Để có thể tra cứu thông tin, người dùng cần cung cấp cho hệ thống một từ khóa liên quan tới thông tin đó. Sau đó hệ thống sẽ tìm kiếm trong nguồn cơ sở dữ liệu hiện tại để đưa ra kết quả tìm kiếm(danh sách thông tin các hôi nghị) cho người dùng |
| Trang web sẽ hiển thị các thông tin hội nghị công nghệ thông tin mới nhất đã được cập nhật | Chức năng cập nhật định kỳ | Để có thể cập nhật thông tin một cách thường xuyên thì sau 1 khoảng thời gian nhất định nào đó, hệ thống sẽ đi tìm kiếm, trích xuất thông tin của một số hội nghị vừa diễn ra. Sau đó thêm vào hệ thống cơ sở dữ liệu của web và cập nhật hiển thị lên giao diện người dùng. |
| Cần biết được danh sách thông tin các hội nghị công nghệ thông tin trên toàn cầu | Chức năng thu thập dữ liệu | Danh sách các trang web và nguồn dữ liệu uy tín về các hội nghị được liệt kê. Crawler sẽ tiến hành thu thập và phân tích mã nguồn, các thành phần trong trang web, lọc các thông tin quan trọng, thông tin trùng lặp và giữ lại các thông tin cơ bản của 1 hội nghị. |
| Trong trường hợp các trang web có khả năng chống crawler hay bị giới hạn về tốc độ truy cập thì phần mềm cần phải xử lý và thông báo tới người dùng | Chức năng xử lý lỗi và ngoại lệ | Hệ thống sẽ phát hiện ra các trang web có chống crawler hay trang web bị giới hạn thời gian truy cập sau đó sẽ thông báo những lỗi tương ứng tới màn hình sử dụng của người dùng |
| Trang web có thể thông báo tới người dùng nếu có một sự kiện, hội nghị vừa mới được diễn ra | Chức năng thông báo và nhắc nhở | Trang web theo định kỳ sẽ cập nhật thêm các thông tin mới nhất về các hội nghị vừa diễn ra(chức năng cập nhật). Sau đó các thông tin này sẽ tổng hợp và gửi thông báo tới người dùng(người đã đặt thông báo cho trang này) để người dùng kịp thời nắm bắt thông tin mới nhất |
| Người dùng có nhu cầu nhận các thông báo mới về các hội nghị sắp diễn ra | Chức năng đăng ký nhận thông báo qua email | Người dùng sẽ cung cấp thông tin về tài khoản email để nhận thông tin về các sự kiện hội nghị sắp diễn ra |
| Người dùng mong muốn có thể xem lại thông tin hội nghị nào đó ở các lần sau | Chức năng lưu trữ | Người dùng sẽ cung cấp thông tin về hội nghị mong muốn được lưu trữ và hệ thống sẽ tiến hành thêm thông tin vào mục lưu trữ để người dùng có thể dễ dàng xem lại sau này |

#### Kiến trúc tổng thể phần mềm

Phần mềm sẽ có một kiến trúc tổng thể như sau:

**1. Phần Thu Thập Dữ Liệu:**

**Scrapy**

**Mô tả:**

Sử dụng Scrapy, một framework Python mạnh mẽ, để thu thập dữ liệu từ các trang web hội nghị.

Scrapy sẽ xử lý việc gửi yêu cầu HTTP, phân tích cú pháp HTML, và trích xuất thông tin cần thiết.

**2. Phần Backend:**

**Django**

**Mô tả:**

Sử dụng Django để xây dựng backend, đảm nhận vai trò trung tâm trong quản lý yêu cầu và tương tác với cơ sở dữ liệu.

Định nghĩa các models trong Django để mô tả cấu trúc dữ liệu của hội nghị và tương tác với cơ sở dữ liệu.

**3. Cơ Sở Dữ Liệu:**

**MySQL**

**Mô tả:**

Dữ liệu thu thập được sẽ được lưu trữ trong một cơ sở dữ liệu SQL, cụ thể là MySQL.

Django sẽ sử dụng ORM để tương tác với cơ sở dữ liệu, đơn giản hóa quá trình lưu trữ và truy xuất dữ liệu.

**4. Giao Diện Người Dùng:**

**React - Typescript:**

**Mô tả:**

Sử dụng React với Typescript để phát triển giao diện người dùng hiện đại và tương tác.

Giao diện người dùng sẽ cung cấp các tính năng hiển thị, tìm kiếm, lọc và xem thông tin về các hội nghị.

**Tổng Hợp:**

Scrapy chịu trách nhiệm về việc thu thập dữ liệu từ các trang web hội nghị và lưu trữ tạm thời.

Django quản lý cơ sở dữ liệu và tương tác với frontend, đồng thời xử lý yêu cầu và cung cấp dữ liệu.

MySQL được sử dụng làm cơ sở dữ liệu chính để lưu trữ thông tin về các hội nghị.

React với Typescript phát triển giao diện người dùng, mang lại trải nghiệm tốt cho người dùng cuối và tương tác linh hoạt với dữ liệu.

#### . Phần cứng

1. Một máy tính có đủ các tài nguyên như bộ vi xử lý, bộ nhớ, ổ cứng để có thể lưu trữ và xử lý dữ liệu thu thập được.

2. Máy tính cần có kết nối internet ổn định: Để Đảm bảo trong quá trình sử dụng phần mềm máy phải có kết nối internet với băng thông đủ và không bị hạn chế truy cập từ phía mạng.

3. Proxy: một số trường hợp crawler sẽ cần phải ẩn danh tính và tránh hạn chế truy cập từ phía các trang web. Proxy sẽ cung cấp một IP giả định cho phép crawler truy cập vào trang web một cách an toàn và không bị chặn.

4. Bộ nhớ: vì phần mềm crawler có thể sẽ yêu cầu một lượng bộ nhớ khá lớn đủ để xử lý và lưu trữ tạm thời các thông tin trong quá trình thu thập. Điều này đặc biệt quan trọng để crawler có thể duyệt và xử lý được nhiều trang web, truy xuất các tệp dữ liệu lớn.

5. Máy Chủ Web:

Cấu Hình Máy Chủ:

* Máy chủ web (Apache/ Nginx) cần được cấu hình để hỗ trợ giao diện người dùng và cung cấp các tệp tĩnh và API từ backend Django.
* Hiệu Suất:
* Cần xác định cấu hình máy chủ để đảm bảo hiệu suất ổn định trong điều kiện tải công việc dự kiến.

6. Máy Chủ Cơ Sở Dữ Liệu:

Loại Cơ Sở Dữ Liệu:

* Sử dụng máy chủ cơ sở dữ liệu SQL ( MySQL ) để lưu trữ và tương tác với dữ liệu từ backend Django.
* Dung Lượng Lưu Trữ:
* Xác định dung lượng cần thiết cho máy chủ cơ sở dữ liệu để đảm bảo tính ổn định và hiệu suất.

7. Máy Chủ Phát Triển và Môi Trường Triển Khai:

* Đủ Tài Nguyên
* Máy chủ phát triển và môi trường triển khai cần có đủ tài nguyên để hỗ trợ quá trình phát triển, kiểm thử, và triển khai ứng dụng một cách hiệu quả.

# Kế hoạch phát triển

#### . Phân tích yêu cầu

* **Công việc:**

+ Từ bài toán sơ lược ban đầu, nhóm sẽ tiến hành thảo luận và đề ra các yêu cầu có thể có đối với website (đứng trên cả góc độ người sử dụng và nhóm phát triển). Từ đó tạo ra một danh sách mô tả yêu cầu, chức năng đi kèm để giải quyết yêu cầu và mô tả chi tiết cho dự án.

* **Thành phẩm cụ thể của giai đoạn:** Danh sách những yêu cầu được đặt ra đứng từ góc độ khách hàng và nhóm phát triển, đi kèm với đó là các chức năng được đề xuất để giải quyết nhu cầu trên và thông tin mô tả chi tiết.
* **Thời gian dự kiến:** 1 tuần

#### . Thiết kế phần mềm

* **Công việc:**
  + Xây dựng sơ đồ cơ sở dữ liệu và mô hình dữ liệu.
  + Thiết kế giao diện người dùng sử dụng công cụ mockup.
  + Dùng UML để mô phỏng các tính năng và thao tác.
* **Thành phẩm cụ thể của giai đoạn:**
  + Kiến trúc tổng thể: Một sơ đồ hoặc một mô hình có chức năng mô tả kiến trúc tổng thể của mô hình Crawler .
  + Thiết kế giao diện người dùng: các bản vẽ màn hình giao diện người dùng.
* **Thời gian dự kiến:** 1 tuần

#### . Cài đặt (implement) phần mềm

* **Công việc:**
  + Phát triển mã nguồn của phần mềm bên phía frontend và backend.
  + Xây dựng và cấu hình cơ sở dữ liệu.
  + Tích hợp các phần tử của phần mềm lại với nhau: web crawler. Cơ sở dữ liệu, giao diện người dùng.
* **Thành phẩm cụ thể của giai đoạn:** Mã nguồn của phần mềm dựa trên thiết kế cơ bản được hoàn thiện.
* **Thời gian dự kiến:** 4 tuần

#### . Kiểm thử phần mềm

* **Công việc:** 
  + Tạo các bộ kiểm tra và kịch bản kiểm thử để kiểm tra chức năng và hiệu suất của phần mềm.
  + Đảm bảo tính bảo mật và an toàn của phần mềm.
  + Sửa lỗi và cải thiện tính năng dựa trên kết quả kiểm thử.
* **Thành phẩm cụ thể của giai đoạn:** 
  + Đảm bảo được tính ổn định và hiệu suất của phần mềm.
  + Phát hiện ra được các lỗi và sự cố nếu có.
* **Thời gian dự kiến:** 3 ngày

#### . Triển khai, bảo trì

* **Công việc:**
  + Triển khai phần mềm lên máy chủ web trên môi trường production.
  + Bảo trì phần mềm, đảm bảo tính ổn định và bảo mật.
  + Cập nhật phần mềm để sửa lỗi (nếu có) và cải tiến tính năng.
* **Thành phẩm cụ thể của giai đoạn:** Phần mềm cơ bản được triển khai thành công trong môi trường thực tế. Tài liệu hướng dẫn chi tiết về cách triển khai và vận hành phần mềm.
* **Thời gian dự kiến:** 3 ngày.
* **Thời gian bảo trì:** vô hạn.

# Kế hoạch nhân sự và chi phí

#### . Kế Hoạch Nhân Sự

* **Nhóm quản lý**
  + Lập kế hoạch, theo dõi tiến độ
  + Khởi tạo dự quán, quản lý mã nguồn
* **Nhóm kỹ thuật** 
  + Lập trình giao diện phía Frontend
  + Lập trình xử lý dữ liệu phía Backend
  + Lập trình thiết kế cài đặt cơ sở dữ liệu
* **Nhóm thiết kế**
  + Thiết kế giao diện người dùng
* **Nhóm Kiểm Thử**
  + Kiểm Thử Phần Mềm Và Đưa ra Đánh Giá

**Bảng Tóm Tắt Kế Hoạch Nhân Sự Cho Dự Án**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 21120505 - Bùi Thị Thanh Ngân | 21120531 - Nguyễn Văn Phúc | 21120540 - Trần Tôn Bửu Quang | 21120600 - Nguyễn Thị Như Ý |
| Nhóm Quản Lý | **X** |  |  | **X** |
| Nhóm Kỹ Thuât |  |  |  |  |
| Chức Năng | **X** | **X** | **X** | **X** |
| Cơ Sở Dữ Liệu |  | **X** | **X** |  |
| Nhóm Thiết Kế | **X** |  |  | **X** |
| Nhóm Kiểm Thử |  | **X** | **X** |  |

#### . Kế Hoạch Chi Phí

* Phí máy chủ web và hosting: 1 triệu VND/ tháng
* Phí duy trì cơ sở dữ liệu: 1 triệu VND/ tháng
* Phí nhân sự: 4 triệu VND/ tháng