**TRƯỜNG ĐẠI HỌC KIẾN TRÚC HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**----- 🙡 🕮 🙣 -----**

**ĐỒ ÁN TỐT NGHIỆP KỸ SƯ**

**NGÀNH: CÔNG NGHỆ THÔNG TIN**

**MÃ NGÀNH: 7480201**

**KHÓA: 2020 - 2025**

**ĐỀ TÀI:**

**NGHIÊN CỨU MỘT SỐ GIẢI PHÁP SỐ HÓA VÀ BẢO MẬT AN TOÀN DỮ LIỆU BÀI GIẢNG KHOA CNTT**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SINH VIÊN THỰC HIỆN:** | | **LƯU TUẤN HUY** | | |
|  | **LỚP:** | **20CN3** |  |  |

**HÀ NỘI 12 - 2024**

**TRƯỜNG ĐẠI HỌC KIẾN TRÚC HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**----- 🙡 🕮 🙣 -----**

**ĐỒ ÁN TỐT NGHIỆP KỸ SƯ**

**NGÀNH: CÔNG NGHỆ THÔNG TIN**

**MÃ NGÀNH: 7480201**

**KHÓA: 2020 - 2025**

**ĐỀ TÀI:**

**NGHIÊN CỨU MỘT SỐ GIẢI PHÁP SỐ HÓA VÀ BẢO MẬT AN TOÀN DỮ LIỆU BÀI GIẢNG KHOA CNTT**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GIẢNG VIÊN HƯỚNG DẪN:** | | **TS. TRẦN TRUNG** | | |
| **SINH VIÊN THỰC HIỆN:** | | **LƯU TUẤN HUY** | | |
|  | **LỚP:** | **20CN3** |  |  |

**HÀ NỘI 12 - 2024**

Mục lục

[**LỜI CẢM ƠN** 5](#_Toc185086774)

[**LỜI MỞ ĐẦU** 6](#_Toc185086775)

[**CHƯƠNG I. CƠ SỞ LÝ THUYẾT** 7](#_Toc185086776)

[**1.1. Mục tiêu và phạm vi nghiên cứu** 7](#_Toc185086777)

[1.1.1. Phương pháp nghiên cứu 8](#_Toc185086778)

[1.1.2. Ý nghĩa thực tiễn 8](#_Toc185086779)

[**1.2. Cơ sở về số hóa dữ liệu** 9](#_Toc185086780)

[1.2.1. Quy trình thực hiện số hóa 9](#_Toc185086781)

[1.2.2. Lợi ích của số hóa 9](#_Toc185086782)

[1.2.3. Các tiêu chuẩn và định dạng số hóa tài liệu: 10](#_Toc185086783)

[**1.3. Cơ sở về bảo mật dữ liệu** 11](#_Toc185086784)

[1.3.1. Một số phương thức tấn công 11](#_Toc185086785)

[1.3.2. Tổng quan về bảo mật dữ liệu 16](#_Toc185086786)

[**1.4. Một số công nghệ và công cụ phổ biến** 17](#_Toc185086787)

[1.4.1. Hệ thống quản lý học tập 17](#_Toc185086788)

[1.4.2. Công cụ mã hóa và quản lý truy cập 19](#_Toc185086789)

[1.4.3. Công cụ sao lưu và bảo vệ dữ liệu 20](#_Toc185086790)

[1.4.4. Công nghệ mới trong bảo mật và số hóa 21](#_Toc185086791)

[**1.5. Các công nghệ sử dụng** 23](#_Toc185086792)

[1.5.1. Tesseract OCR 23](#_Toc185086793)

[1.5.2. Node.js và MySQL 24](#_Toc185086794)

[**CHƯƠNG II. PHÂN TÍCH THIẾT KẾ HỆ THỐNG** 27](#_Toc185086795)

[**2.1 Phân tích yêu cầu** 27](#_Toc185086796)

[2.1.1 Mô tả nghiệp vụ 27](#_Toc185086797)

[2.1.2 Yêu cầu chức năng 28](#_Toc185086798)

[2.1.3 Yêu cầu phi chức năng 28](#_Toc185086799)

[**2.2 Phân tích hệ thống** 28](#_Toc185086800)

[2.2.1 Xác định các tác nhân chính của hệ thống 28](#_Toc185086801)

[2.2.2 Các chức năng của tác nhân 29](#_Toc185086802)

[2.2.3 Biểu đồ Use Case tổng quát 30](#_Toc185086803)

[2.2.4 Biều đồ chức năng 30](#_Toc185086804)

[**2.3 Thiết kế hệ thống** 31](#_Toc185086805)

[2.3.1 Thiết kế cơ sở dữ liệu 31](#_Toc185086806)

[2.3.2 Thiết kế các giao diện 34](#_Toc185086807)

[**CHƯƠNG III. XÂY DỰNG VÀ TRIỂN KHAI HỆ THỐNG** 36](#_Toc185086808)

[**3.1 Xây dựng các chức năng** 36](#_Toc185086809)

[3.1.1 Cấu trúc chương trình 36](#_Toc185086810)

[3.1.2 Các đối tượng và chức năng tương ứng 36](#_Toc185086811)

[**3.2 Kiểm thử các chức năng** 38](#_Toc185086812)

[3.2.1 Chức năng đăng nhập 38](#_Toc185086813)

[3.2.2 Chức năng tải tài liệu 39](#_Toc185086814)

[3.2.3 Chức năng tìm kiếm 39](#_Toc185086815)

[**KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** 40](#_Toc185086816)

[**Tổng kết đánh giá** 40](#_Toc185086817)

[**Hướng phát triển** 40](#_Toc185086818)

[**TÀI LIỆU THAM KHẢO** 41](#_Toc185086819)

# **LỜI CẢM ƠN**

Em xin gửi lời cảm ơn chân thành nhất tới quý thầy cô Trường Đại Học Kiến Trúc Hà Nội, những người đã dìu dắt, dạy dỗ em tận tình, và truyền đạt những kiến thức và nhiều bài học quý giá xuyên suốt quãng thời gian em theo học tại trường.

Em xin trân trọng gửi lời cảm ơn đến tất cả quý thầy cô trong khoa Công Nghệ Thông Tin, đặc biệt là giảng viên TS. Trần Trung, người đã tận tình hướng dẫn góp ý và chỉ bảo em trong suốt thời gian làm đồ án tốt nghiệp. Với sự giúp đỡ của thầy, em đã có được những định hướng trong nghiên cứu cũng như triển khai thực hiện các yêu cầu trong quá trình làm đồ án tốt nghiệp.

Ngoài ra em cũng xin cảm ơn tới những người thân và gia đình đã quan tâm, động viên và luôn tạo những điều kiện thuận lợi nhất trong quá trình học và làm tốt nghiệp. Bên cạnh đó không thể thiếu là các bạn trong tập thể lớp 20C3 đã luôn đoàn kết gắn bó trong các hoạt động và việc học tập, giúp đỡ em rất nhiều trong những năm tháng qua.

Em xin chân thành cảm ơn!

# **LỜI MỞ ĐẦU**

Trong bối cảnh chuyển đổi số ngày càng phát triển, giáo dục là một trong những lĩnh vực chịu ảnh hưởng mạnh mẽ nhất. Việc số hóa tài liệu giảng dạy không chỉ giúp nâng cao hiệu quả giảng dạy, tiết kiệm thời gian, mà còn tạo điều kiện cho việc truy cập, chia sẻ và lưu trữ dữ liệu bài giảng một cách thuận tiện. Tuy nhiên, quá trình này cũng đặt ra nhiều thách thức, đặc biệt là về mặt bảo mật dữ liệu, khi thông tin nhạy cảm có nguy cơ bị truy cập trái phép, rò rỉ hoặc bị tấn công mạng.

Đối với Khoa Công nghệ Thông tin, nơi lưu trữ và giảng dạy nhiều nội dung liên quan đến lập trình, cơ sở dữ liệu và an toàn thông tin, việc đảm bảo tính bảo mật cho dữ liệu bài giảng là một yêu cầu cấp thiết. Những lỗ hổng bảo mật hoặc việc không kiểm soát tốt quyền truy cập có thể dẫn đến các hậu quả nghiêm trọng như đánh cắp thông tin, sửa đổi nội dung bài giảng hoặc làm gián đoạn quá trình học tập và giảng dạy.

Mục tiêu của đề tài là: Nghiên cứu một số giải pháp số hóa và bảo mật an toàn dữ liệu bài giảng khoa CNTT. Nội dung có cấu trúc gồm các mục như sau:

Chương 1: Cơ sở lý thuyết: Xác định mục tiêu và phạm vi, các nội dung cơ bản về số hóa dữ liệu, bảo mật dữ liệu, giới thiệu một số công cụ và công nghệ được sử dụng.

Chương 2: Phân tích và thiết kế hệ thống: Tiến hành xác định và phân tích yêu cầu về chức năng, vẽ các sơ đồ cơ bản phục vụ cho hệ thống.

Chương 3: Xây dựng và triển khai hệ thống: Ứng dụng các giải pháp và biện pháp thực tế, lấy các ví dụ và triển khai trên nền tảng website.

Chương 4: Kết luận và hướng phát triển: Đưa ra kết luận chung và phương hướng phát triển sau này.

Bài báo cáo đồ án tốt nghiệp của em vẫn còn nhiều hạn chế và thiếu sót, rất mong được quý thầy cô góp ý chỉ bảo để giúp báo cáo hoàn thiện hơn.

Em xin chân thành cảm ơn!

# **CHƯƠNG I. CƠ SỞ LÝ THUYẾT**

## **1.1. Mục tiêu và phạm vi nghiên cứu**

Đề tài hướng đến các mục tiêu sau:

* Nghiên cứu và lựa chọn giải pháp số hóa dữ liệu bài giảng: Phân tích và áp dụng các công cụ, nền tảng hiện có (ví dụ: Moodle, Google Classroom) để chuyển đổi tài liệu từ định dạng truyền thống sang định dạng số một cách hiệu quả.
* Đảm bảo an toàn dữ liệu bài giảng: Đề xuất và triển khai các biện pháp bảo mật như mã hóa dữ liệu, quản lý phân quyền truy cập, và sao lưu tự động nhằm giảm thiểu nguy cơ rò rỉ hoặc mất mát dữ liệu.
* Ứng dụng thực tiễn: Sau khi nghiên cứu, tiến hành triển khai một hệ thống minh họa trên nền tảng website gồm một số tính năng: Phân quyền người dùng, đăng tải tài liệu, mã hóa và lưu trữ an toàn và linh hoạt.
* Đánh giá hiệu quả của các giải pháp sử dụng: Phân tích mức độ phù hợp, tính bảo mật và khả năng đáp ứng nhu cầu thực tế của các công cụ đã chọn.

Đối tượng nghiên cứu: Tài liệu bài giảng của Khoa CNTT, bao gồm:

* Tài liệu học tập (slide, PDF, video).
* Đề kiểm tra, bài tập.
* Các nội dung học thuật liên quan.

Phạm vi nghiên cứu:

* Về số hóa dữ liệu: Tích hợp và thử nghiệm các nền tảng quản lý học tập như Moodle, Google Classroom, Microsoft Teams để số hóa tài liệu giảng dạy.
* Về bảo mật dữ liệu: Tập trung vào các giải pháp mã hóa dữ liệu, quản lý phân quyền truy cập và sao lưu, sử dụng các công cụ phổ biến như VeraCrypt hoặc Auth0.
* Ứng dụng: Xây dựng một nền tảng website để minh họa các tính năng đã nghiên cứu và các phương pháp được chọn lựa.

Hạn chế nghiên cứu:

* Không phát triển hệ thống bảo mật, số hóa từ ban đầu, chỉ tập trung thử nghiệm và đánh giá các công cụ, giải pháp sẵn có.
* Chưa nghiên cứu toàn diện về bảo mật toàn bộ hệ thống CNTT của Khoa, chỉ giới hạn trong phạm vi dữ liệu bài giảng.

### 1.1.1. Phương pháp nghiên cứu

Nghiên cứu tài liệu:

* Thu thập và phân tích các tài liệu liên quan đến số hóa tài liệu giảng dạy và bảo mật dữ liệu.
* Đánh giá tính năng, ưu điểm, nhược điểm của các công cụ số hóa và bảo mật hiện có.

Phân tích và thử nghiệm giải pháp:

* Cài đặt và thử nghiệm các công cụ số hóa dữ liệu như Moodle, Google Classroom.
* Triển khai thử nghiệm các biện pháp bảo mật như mã hóa dữ liệu, phân quyền truy cập.
* Xây dựng nền tảng website minh họa với Node.js và MySQL, tích hợp Tesseract OCR để xử lý tài liệu số hóa.

Đánh giá và so sánh:

* Đánh giá hiệu quả của hệ thống dựa trên các tiêu chí:
* Tính thân thiện với người dùng: Giao diện dễ sử dụng, khả năng truy cập nhanh chóng.
* Hiệu suất hoạt động: Tốc độ truy cập, khả năng quản lý dữ liệu.
* Mức độ an toàn: Khả năng chống lại các nguy cơ rò rỉ, tấn công dữ liệu.
* So sánh các công cụ để đề xuất giải pháp tối ưu nhất cho Khoa CNTT.

### 1.1.2. Ý nghĩa thực tiễn

Đối với Khoa CNTT:

* Đảm bảo dữ liệu bài giảng được lưu trữ và truy cập an toàn, giúp nâng cao chất lượng giảng dạy.
* Tăng tính chuyên nghiệp và hiện đại hóa trong việc quản lý tài liệu học thuật.
* Cung cấp mô hình hệ thống minh họa thực tế có thể áp dụng hoặc mở rộng trong tương lai

Đối với sinh viên và giảng viên:

* Tạo điều kiện thuận lợi trong việc truy cập tài liệu học tập mọi lúc, mọi nơi.
* Giảm thiểu mất mát dữ liệu, đảm bảo sự tin tưởng trong sử dụng tài liệu số.
* Hỗ trợ hiệu quả trong quy trình học tập và giảng dạy thông qua các ví dụ thực tế như tải lên các tài liệu, mã hóa và lưu trữ an toàn trên nền tảng website.

## **1.2. Cơ sở về số hóa dữ liệu**

Số hóa dữ liệu giáo dục bao gồm việc chuyển đổi nội dung giảng dạy từ các tài liệu truyền thống (sách, giấy tờ) sang định dạng số, như file PDF, eBook, video hoặc nội dung tương tác trên nền tảng học tập trực tuyến.

### 1.2.1. Quy trình thực hiện số hóa

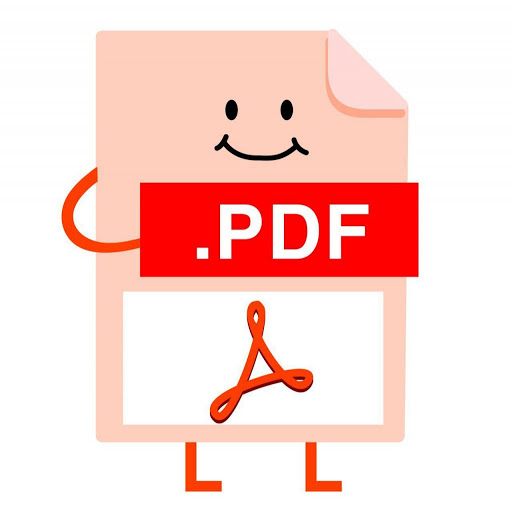
* Thu thập tài liệu: Tập hợp tài liệu giảng dạy, giáo trình và các tài liệu nghiên cứu cần số hóa.
* Chuyển đổi định dạng: Sử dụng các công cụ chuyển đổi dữ liệu từ giấy hoặc các định dạng khác khác sang cái file số hóa (PDF, ePub, HTML5, video).
* Tổ chức lưu trữ: Lưu trữ tài liệu trên các nền tảng thư viện số hoặc trên các hệ thống quản lý học tập để dễ dàng quản lý, sử dụng và tra cứu.
* Tổ chức quản lý: Phân loại các tài liệu rõ ràng, áp dụng các biện pháp bảo mật để tránh mất mát hoặc truy cập trái phép.

### 1.2.2. Lợi ích của số hóa

* Hiệu quả trong quản lý: Giảm thời gian và công sức quản lý tài liệu.
* Tổ chức khoa học: Dễ dàng phân loại, tìm kiếm tài liệu nhanh chóng.
* Tiết kiếm không gian: Không cần không gian lưu trữ vật lý, tận dụng các dịch vụ lưu trữ đám mây
* Tiện lợi trong chia sẻ và truy cập:
  + Học viên và giảng viên có thể truy cập tài liệu từ mọi nơi, mọi lúc.
  + Hỗ trợ học tập từ xa và mô hình giáo dục linh hoạt.
* Tránh nguy cơ mất mát tài liệu vật lý do hỏng hóc hoặc thất lạc.

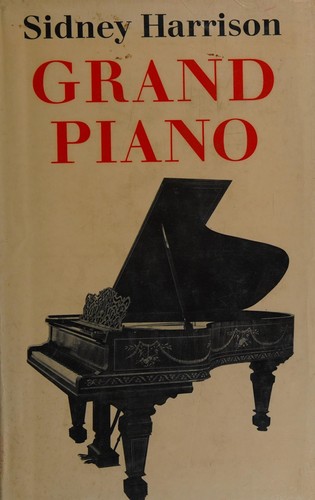
### 1.2.3. Các tiêu chuẩn và định dạng số hóa tài liệu:

PDF:



* Định dạng phổ biến, hỗ trợ bảo mật bằng mật khẩu và không thay đổi bố cục tài liệu trên các thiết bị khác nhau.
* Sử dụng trong lưu trữ bài giảng, tài liệu nghiên cứu.

ePub:



* Định dạng tối ưu cho sách điện tử
* Hỗ trợ nhiều kích thước màn hình, phù hợp với thiết bị di động.

HTML5:



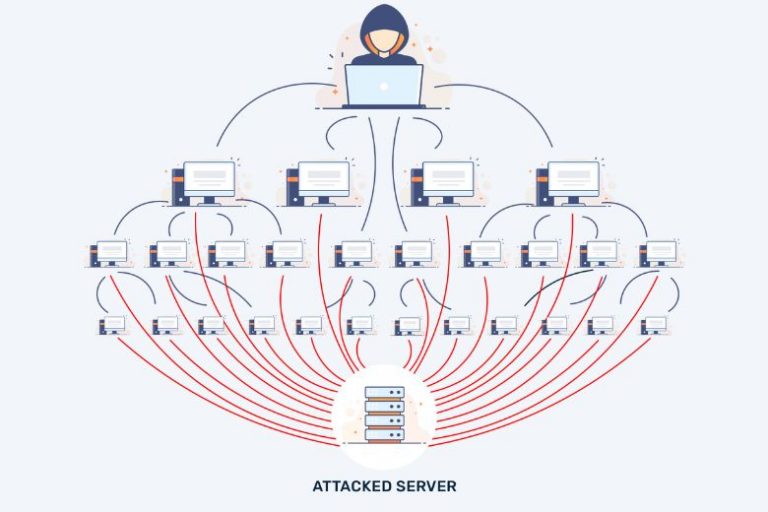
* Hỗ trợ xây dựng tài liệu tương tác với hình ảnh động, video, và bài tập trực tuyến.
* Ứng dụng trong các khóa học eLearning.

## **1.3. Cơ sở về bảo mật dữ liệu**

### ****1.3.1. Một số phương thức tấn công****

Trong bối cảnh số hóa và sử dụng công nghệ trong giảng dạy, dữ liệu bài giảng dễ trở thành mục tiêu của các cuộc tấn công mạng. Các phương thức tấn công phổ biến bao gồm:

**Tấn công từ chối dịch vụ (DDoS)**



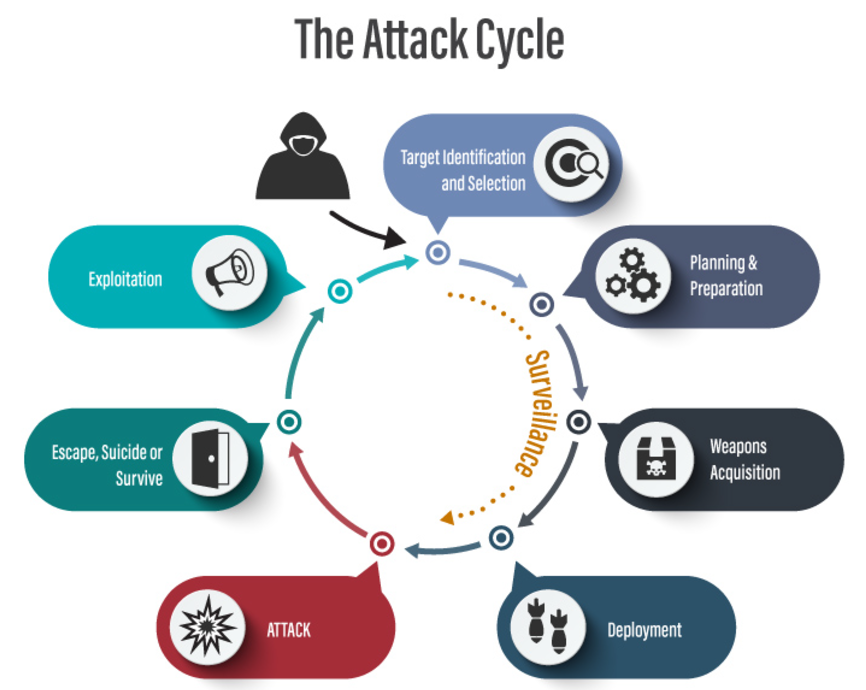
* **Mô tả**: Kẻ tấn công làm quá tải hệ thống quản lý học tập (LMS) hoặc máy chủ lưu trữ dữ liệu bài giảng bằng cách gửi lượng lớn yêu cầu đồng thời, làm hệ thống không thể phục vụ người dùng hợp pháp.
* **Tác động**:
  + Làm gián đoạn hoạt động học tập và giảng dạy.
  + Ảnh hưởng đến trải nghiệm người dùng.
  + Giảm uy tín của hệ thống học tập.
* **Ví dụ**: Một hệ thống LMS như Moodle bị tấn công DDoS, khiến sinh viên không thể truy cập tài liệu bài giảng.

**Tấn công bằng mã độc (Malware)**



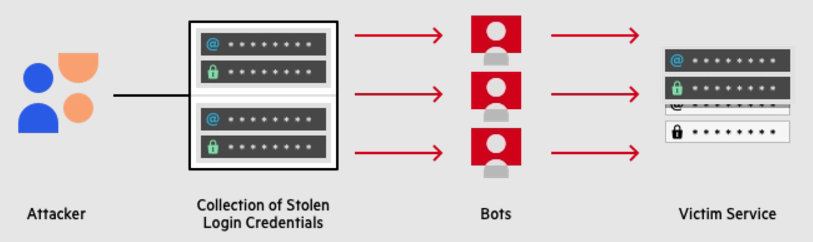
* **Mô tả**: Kẻ tấn công cài đặt phần mềm độc hại vào hệ thống để đánh cắp, sửa đổi hoặc phá hủy dữ liệu bài giảng.
* **Hình thức mã độc phổ biến**:
  + **Ransomware**: Mã hóa dữ liệu và yêu cầu tiền chuộc để khôi phục.
  + **Spyware**: Theo dõi và thu thập thông tin từ hệ thống.
  + **Trojan**: Mở cổng hậu để truy cập trái phép vào dữ liệu.
* **Tác động**:
  + Làm mất dữ liệu bài giảng.
  + Gây gián đoạn quá trình giảng dạy.
  + Tăng chi phí khắc phục và nâng cấp hệ thống.

**Tấn công khai thác lỗ hổng bảo mật (Vulnerability Exploitation)**



* **Mô tả**: Tận dụng lỗ hổng trong phần mềm hoặc hệ thống quản lý để truy cập trái phép, sửa đổi hoặc xóa dữ liệu bài giảng.
* **Ví dụ**:
  + Lỗi SQL Injection: Kẻ tấn công sử dụng truy vấn SQL độc hại để truy cập cơ sở dữ liệu chứa bài giảng.
  + Lỗi XSS (Cross-Site Scripting): Kẻ tấn công chèn mã độc vào các trang web liên quan đến LMS.
* **Tác động**:
  + Đánh cắp dữ liệu nhạy cảm (nội dung bài giảng, đề thi).
  + Làm thay đổi nội dung giảng dạy.

**Tấn công qua tài khoản người dùng (Credential Attacks)**

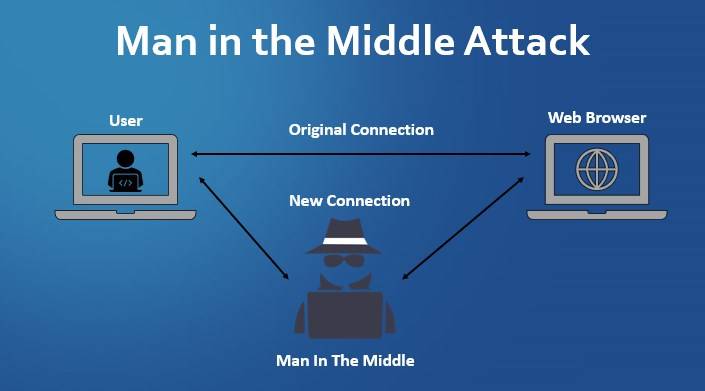


* **Mô tả**: Kẻ tấn công cố gắng đánh cắp hoặc sử dụng trái phép tài khoản của giảng viên hoặc quản trị viên.
* **Phương pháp**:
  + **Phishing**: Lừa người dùng cung cấp thông tin đăng nhập thông qua email hoặc trang web giả mạo.
  + **Brute Force**: Thử mọi tổ hợp mật khẩu để truy cập tài khoản.
  + **Credential Stuffing**: Sử dụng thông tin đăng nhập bị rò rỉ từ hệ thống khác để tấn công.
* **Tác động**:
  + Truy cập và sửa đổi nội dung bài giảng.
  + Tạo tài liệu giả mạo hoặc phát tán thông tin sai lệch.

**Tấn công nội bộ (Insider Threats)**

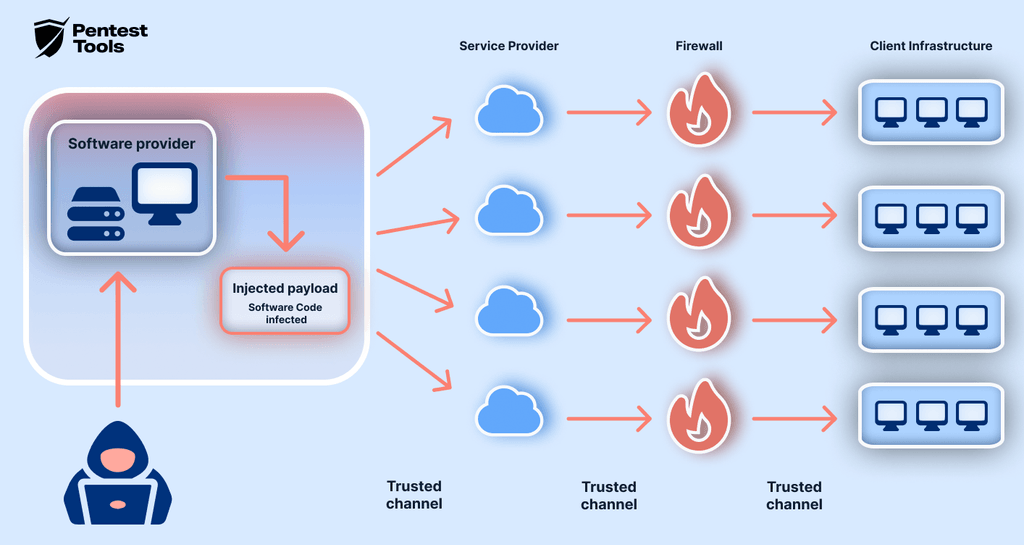
* **Mô tả**: Người bên trong tổ chức, như giảng viên hoặc nhân viên, sử dụng quyền truy cập hợp pháp để đánh cắp hoặc phá hoại dữ liệu bài giảng.
* **Tác động**:
  + Dữ liệu bài giảng bị rò rỉ ra bên ngoài.
  + Gây tổn thất về tài chính và uy tín.

**Tấn công mạng lưới (Man-in-the-Middle - MITM)**



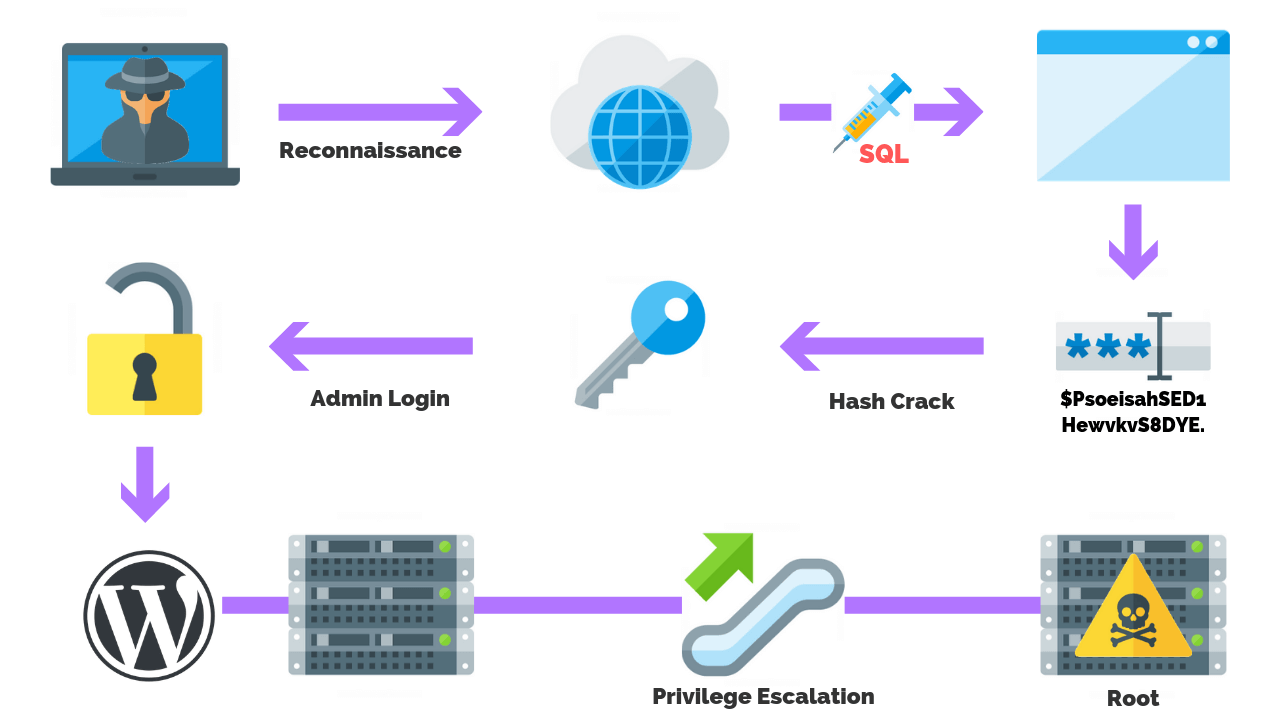
* **Mô tả**: Kẻ tấn công chặn luồng dữ liệu giữa giảng viên/sinh viên và hệ thống LMS để đánh cắp hoặc thay đổi nội dung.
* **Ví dụ**:
  + Chặn các tệp bài giảng tải xuống và chèn mã độc.
  + Đánh cắp thông tin đăng nhập trong quá trình người dùng truy cập hệ thống.
* **Tác động**:
  + Mất dữ liệu nhạy cảm.
  + Gây gián đoạn và mất uy tín của hệ thống.

**Tấn công thông qua chuỗi cung ứng (Supply Chain Attacks)**



* **Mô tả**: Kẻ tấn công nhắm vào các công cụ hoặc phần mềm hỗ trợ số hóa bài giảng để chèn mã độc hoặc tạo lỗ hổng.
* **Ví dụ**: Một plugin trên Moodle bị cài mã độc làm rò rỉ dữ liệu.
* **Tác động**:
  + Ảnh hưởng đến toàn bộ hệ thống giảng dạy.
  + Đặt người dùng vào nguy cơ đánh cắp dữ liệu nhạy cảm.

**Tấn công từ chối phân quyền (Privilege Escalation)**



* **Mô tả**: Kẻ tấn công lợi dụng lỗ hổng trong phân quyền để nâng cấp quyền truy cập từ sinh viên thành quản trị viên.
* **Tác động**:
  + Làm thay đổi dữ liệu bài giảng.
  + Xóa hoặc làm mất tài liệu quan trọng.

### 1.3.2. Tổng quan về bảo mật dữ liệu

Bảo mật dữ liệu là việc bảo vệ thông tin khỏi các mối đe dọa như truy cập trái phép, rò rỉ, hoặc sửa đổi không mong muốn.

Nguyên tắc cơ bản:

* Bảo mật (Confidentiality): Chỉ người được ủy quyền mới có quyền truy cập dữ liệu.
* Toàn vẹn (Integrity): Dữ liệu không bị thay đổi hoặc hỏng hóc ngoài ý muốn.
* Khả dụng (Availability): Đảm bảo dữ liệu luôn sẵn sàng khi cần thiết.

Phân tích các nguy cơ bảo mật đối với dữ liệu giáo dục:

* Xâm nhập trái phép: Người không được cấp quyền cố tình truy cập vào dữ liệu bài giảng.
* Ví dụ: Truy cập trái phép vào hệ thống LMS hoặc tài khoản lưu trữ đám mây.
* Rò rỉ dữ liệu: Thông tin bài giảng, đề thi bị lộ ra ngoài, gây ảnh hưởng đến chất lượng học tập và bảo mật học thuật.
* Mã độc và tấn công mạng: Các phần mềm độc hại hoặc mã hóa dữ liệu.
* Ví dụ: Tấn công từ chối dịch vụ (DDoS) làm gián đoạn hệ thống LMS.

Các biện pháp bảo mật phổ biến:

Mã hóa dữ liệu:

* Mã hóa đối xứng (AES): Một khóa được sử dụng để mã hóa và giải mã.
* Mã hóa bất đối xứng (RSA): Sử dụng hai khóa (công khai và riêng tư) để bảo mật thông tin.

Quản lý phân quyền truy cập:

* Cấp quyền theo vai trò (Role-Based Access Control - RBAC), ví dụ: Giảng viên có quyền chỉnh sửa nội dung, sinh viên chỉ được xem.

Sao lưu dữ liệu:

* Sao lưu định kỳ trên các nền tảng đám mây hoặc ổ cứng cục bộ để khôi phục khi xảy ra sự cố.

## **1.4. Một số công nghệ và công cụ phổ biến**

### 1.4.1. Hệ thống quản lý học tập

Moodle:



* Mã nguồn mở, cho phép tùy chỉnh theo nhu cầu.
* Hỗ trợ tính năng quản lý khóa học, bài kiểm tra, và giao tiếp giữa giảng viên và sinh viên.
* Có khả năng tích hợp nhiều plugin nâng cao.

Google Classroom:



* Miễn phí, triển khai dễ dàng, đặc biệt phù hợp với môi trường giáo dục phổ thông.
* Tích hợp tốt với các dịch vụ khác của Google như Drive, Docs, Google Meet.
* Giao diện thân thiện, phù hợp với người dùng.

Microsoft Teams:



* Tích hợp với bộ công cụ Office 365, hỗ trợ soạn thảo tài liệu, quản lý các bài giảng, và khởi chạy các lớp học trực tuyến.
* Tích hợp tính năng lưu trữ video và các văn bản tài liệu, hỗ trợ làm việc theo nhóm thuận lợi trong việc giáo dục và đào tạo.

### 1.4.2. Công cụ mã hóa và quản lý truy cập

VeraCrypt:



* Công cụ mã nguồn mở, hộ trợ tốt trong việc mã hóa các tệp tin, ổ đĩa và các thư mục quan trọng khác.
* Được sử dụng trong bảo mật tài liệu về bài giảng, hạn chế trong việc mất mát dữ liệu khi bị tấn công.

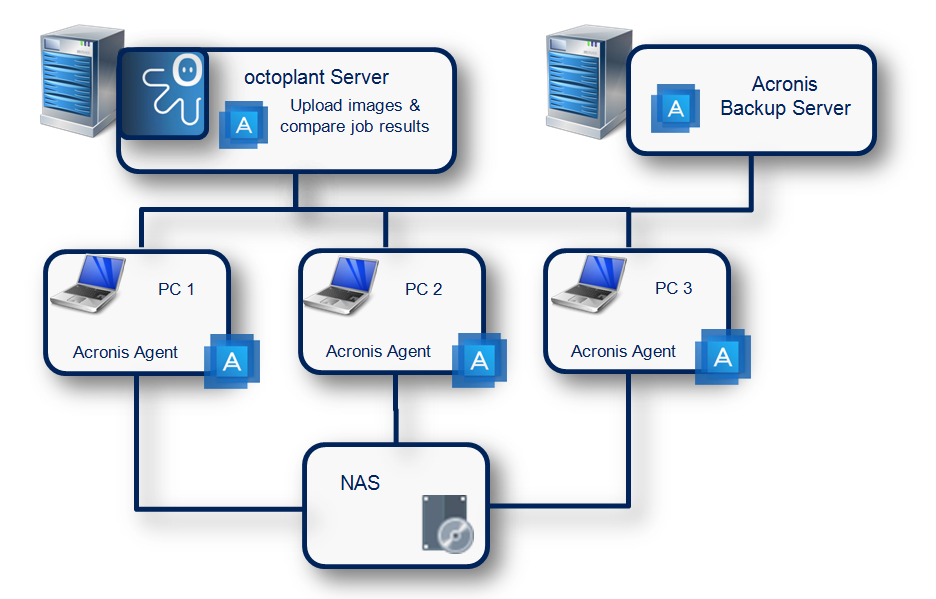
Auth0:



* Nền tảng hỗ trợ tính năng xác thực mạnh mẽ.
* Hỗ trợ các phương thức đăng nhập an toàn: Xác thực đa yếu tố Multi-factor Authentication (MFA) và xác thực một lần Single sign-on (SSO).
* Dễ dàng tích hợp với các phần mềm quản lý học tập khác.

### 1.4.3. Công cụ sao lưu và bảo vệ dữ liệu

Acronis Backup:



* Sao lưu và khôi phục dữ liệu nhanh chóng.
* Tích hợp các tính năng bảo mật.
* Hỗ trợ lưu trữ đám mây và quản lý từ xa.

Google Drive:



* Lưu trữ miễn phí với dung lượng 15GB, có thể mở rộng qua các gói trả phí khác.
* Tiện lợi trong việc chỉnh sửa và chia sẻ tài liệu.
* Tích hợp được với nhiều công cụ, hỗ trợ tốt trong việc quản lý.

OneDirve:



* Hỗ trợ 5GB dung lượng miễn phí.
* Tích hợp với bộ công cụ của Microsoft Office.
* Hỗ trợ mã hóa tài liệu, có thể khôi phục dữ liệu sau các cuộc tấn công mã độc.

### 1.4.4. Công nghệ mới trong bảo mật và số hóa

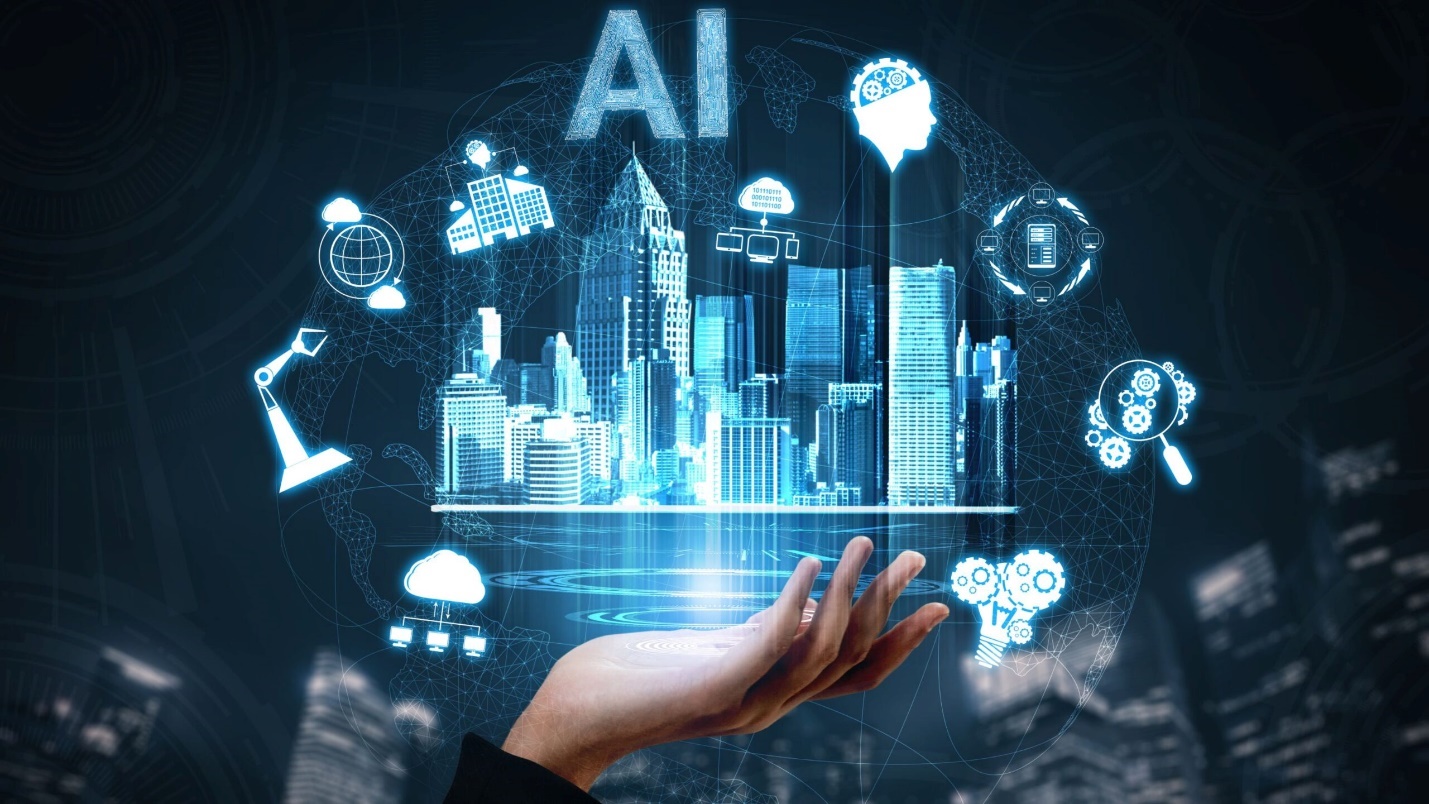
Blockchain:



Ứng dụng trong giáo dục:

* Lưu trữ và xác minh dữ liệu học tập:
  + Đảm bảo tính toàn vẹn và bất biến của dữ liệu.
  + Ví dụ các bằng khen, chứng chỉ và các bảng điểm của sinh viên được lưu trên mạng lưới của blockchain, giúp xác minh nhanh chóng và tránh việc giả mạo.
* Bạo mật dữ liệu trong giảng dạy:
  + Đảm bảo nội dung không bị xóa, chỉnh sửa hoặc truy cập trái phép.
* Lợi ích:
  + Tăng cường tính an toàn và bảo mật dữ liệu.
  + Dễ dàng theo dõi các lịch sử truy cập và sửa đổi.

Trí tuệ nhân tạo Artificial Intelligence (AI):



Ứng dụng trong bảo mật:

* Phát hiện các truy cập bất thường: AI có thể phân tích và phát hiện các hành vi bất thường, đưa ra các cảnh báo sớm về nguy cơ xảy ra các cuộc tấn công.
* Hỗ trợ tự động hóa dữ liệu: Hạn chế nguy cơ rò rỉ dữ liệu, hỗ trợ mã hóa những tài liệu quan trọng và nhạy cảm.

Ứng dụng trong số hóa:

* Trích xuất dữ liệu tự động: AI hỗ trợ chuyển đổi tài liệu sang định dạng số, từ đó giúp tiết kiệm thời gian và công sức.
* Nâng cao tính tương tác và hiệu quả học tập: Gợi ý điều chỉnh bài giảng, hỗ trợ chuyển đổi các nội dung.

## **1.5. Các công nghệ sử dụng**

### 1.5.1. Tesseract OCR

Tesseract OCR là một phần mềm mã nguồn mở được sử dụng để chuyển đổi hình ảnh chứa văn bản thành dữ liệu văn bản có thể chỉnh sửa được. Được phát triển lần đầu bởi HP và hiện được duy trì bởi Google, Tesseract đã trở thành một trong những công cụ OCR mạnh mẽ nhất hiện nay, với khả năng nhận diện văn bản từ nhiều ngôn ngữ khác nhau, bao gồm cả tiếng Việt.

**Vai trò của Tesseract OCR trong số hóa dữ liệu:**

* **Chuyển đổi tài liệu giấy thành dữ liệu số:** Tesseract OCR giúp số hóa tài liệu giấy bằng cách nhận diện và chuyển đổi văn bản từ hình ảnh hoặc tài liệu quét (scan) thành văn bản có thể chỉnh sửa. Điều này giúp loại bỏ sự phụ thuộc vào giấy tờ, tiết kiệm thời gian, và dễ dàng quản lý và tìm kiếm tài liệu.
* **Tính chính xác và độ tin cậy cao:** Với sự cải tiến qua nhiều phiên bản, Tesseract ngày càng có độ chính xác cao trong việc nhận diện văn bản từ các tài liệu có chất lượng thấp hoặc phức tạp, như tài liệu in nghiêng, chữ viết tay, hay các hình ảnh mờ. Điều này đặc biệt hữu ích trong các hệ thống số hóa tài liệu giáo dục, nơi chất lượng tài liệu có thể thay đổi.
* **Ứng dụng trong giáo dục:** Trong lĩnh vực giáo dục, Tesseract OCR có thể giúp số hóa tài liệu giảng dạy, sách giáo khoa, bài kiểm tra, hay các bảng điểm. Các tài liệu này có thể được lưu trữ và truy xuất dễ dàng hơn, phục vụ cho việc nghiên cứu và giảng dạy. Tesseract còn giúp tự động hóa quá trình nhập liệu, giảm thiểu sai sót do nhập tay.
* **Hỗ trợ nhiều ngôn ngữ và định dạng văn bản:** Tesseract hỗ trợ nhận diện văn bản từ hơn 100 ngôn ngữ và có thể tùy chỉnh để nhận diện các loại chữ viết đặc biệt hoặc các phông chữ không chuẩn. Điều này mang lại sự linh hoạt cao khi áp dụng vào các môi trường học thuật đa dạng với các tài liệu viết bằng nhiều ngôn ngữ khác nhau.
* **Tích hợp với các ứng dụng số hóa và quản lý tài liệu:** Tesseract OCR có thể dễ dàng tích hợp vào các hệ thống quản lý tài liệu và học tập (LMS) để hỗ trợ tự động số hóa và tìm kiếm tài liệu. Việc tích hợp với các công nghệ web như Node.js và cơ sở dữ liệu MySQL có thể giúp lưu trữ và truy xuất dữ liệu OCR một cách hiệu quả, cho phép giảng viên và sinh viên dễ dàng tìm kiếm tài liệu giảng dạy, sách giáo khoa, hay tài liệu nghiên cứu.
* **Tự động hóa và tiết kiệm thời gian:** Việc sử dụng Tesseract OCR trong số hóa tài liệu giúp tự động hóa quy trình số hóa thay vì phải nhập tay từng dữ liệu. Điều này không chỉ tiết kiệm thời gian mà còn giúp giảm thiểu lỗi do con người khi làm việc với lượng tài liệu lớn.

Tesseract OCR đóng vai trò quan trọng trong quá trình số hóa dữ liệu, đặc biệt trong giáo dục. Việc tích hợp Tesseract OCR vào các hệ thống học tập trực tuyến giúp cải thiện hiệu quả công việc và đảm bảo việc lưu trữ, bảo mật và truy cập dữ liệu dễ dàng hơn.

### 1.5.2. Node.js và MySQL

Node.js là một môi trường runtime JavaScript mở, sử dụng engine V8 của Google Chrome để thực thi mã JavaScript. Nó cho phép các nhà phát triển viết ứng dụng máy chủ bằng JavaScript, giúp việc phát triển ứng dụng web trở nên dễ dàng hơn, vì JavaScript vốn đã được sử dụng chủ yếu ở phía client trong các ứng dụng web. Node.js được thiết kế để tối ưu cho các ứng dụng có thể xử lý nhiều kết nối đồng thời, như các dịch vụ web, API, và ứng dụng thời gian thực.

Các đặc điểm chính của Node.js:

* Hỗ trợ xử lý bất đồng bộ và hướng sự kiện: Node.js sử dụng mô hình xử lý bất đồng bộ, nghĩa là nó không chặn các tác vụ (chẳng hạn như đọc file hay truy vấn cơ sở dữ liệu). Điều này giúp ứng dụng có thể xử lý nhiều yêu cầu đồng thời mà không bị gián đoạn.
* Hiệu suất cao: Node.js có hiệu suất rất cao, đặc biệt khi xử lý các tác vụ Input/Output (nhập/xuất) như truy vấn cơ sở dữ liệu, đọc file.
* Sử dụng JavaScript ở cả hai phía client và server: Điều này giúp các nhà phát triển có thể sử dụng một ngôn ngữ chung trong toàn bộ dự án, giảm thiểu sự phức tạp và dễ dàng phối hợp giữa các phần của ứng dụng.

Node.js rất phù hợp với các ứng dụng cần khả năng xử lý nhiều kết nối đồng thời, ví dụ như API web, dịch vụ thời gian thực, hoặc ứng dụng đa người dùng. Khi kết hợp với các công cụ và thư viện mã nguồn mở Node Package Manager (npm), Node.js cung cấp một hệ sinh thái rộng lớn để phát triển nhanh chóng các ứng dụng web.

MySQL là một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở, được sử dụng phổ biến trong các ứng dụng web. Nó hỗ trợ quản lý và lưu trữ dữ liệu dưới dạng bảng với các mối quan hệ giữa các bảng này. MySQL rất được ưa chuộng nhờ vào khả năng mở rộng, tính ổn định và hiệu suất cao, cùng với việc dễ dàng tích hợp với nhiều ngôn ngữ lập trình và môi trường phát triển khác nhau.

Các đặc điểm chính của MySQL:

* Môi trường SQL mạnh mẽ: MySQL hỗ trợ ngôn ngữ truy vấn cấu trúc (SQL), giúp dễ dàng thực hiện các thao tác như chèn, cập nhật, xóa, và truy vấn dữ liệu.
* Quản lý giao dịch mạnh mẽ: MySQL hỗ trợ giao dịch ACID, đảm bảo tính toàn vẹn dữ liệu trong các hệ thống yêu cầu xử lý dữ liệu đáng tin cậy.
* Tính mở rộng cao: MySQL có thể được sử dụng trong các ứng dụng nhỏ lẫn lớn, với khả năng mở rộng để đáp ứng nhu cầu lưu trữ và truy vấn dữ liệu khối lượng lớn.

Tích hợp Node.js với MySQL:

Khi kết hợp Node.js với MySQL, Node.js có thể giúp xây dựng các API backend mạnh mẽ, nơi các dữ liệu được lưu trữ và truy xuất từ MySQL. Điều này rất hữu ích khi cần phát triển các hệ thống quản lý dữ liệu, như trong ứng dụng số hóa tài liệu hoặc quản lý học tập.

Các thư viện như mysql hoặc sequelize trong Node.js giúp kết nối và thao tác với cơ sở dữ liệu MySQL một cách dễ dàng và hiệu quả. Một số ứng dụng điển hình trong việc tích hợp này là:

* Quản lý người dùng: Lưu trữ thông tin người dùng, đăng nhập, phân quyền trong các ứng dụng như hệ thống quản lý học tập (LMS).
* Xử lý file: Lưu trữ thông tin về các file tài liệu, hỗ trợ tải lên và tải xuống các tài liệu số hóa.

Kết hợp Node.js, MySQL và Tesseract OCR trong ứng dụng web:

* Node.js sẽ xử lý các yêu cầu HTTP từ người dùng (như tải tài liệu lên).
* Tesseract OCR sẽ được tích hợp vào ứng dụng để xử lý các tài liệu hình ảnh (chẳng hạn như các tài liệu quét hoặc hình ảnh chứa văn bản).
* Các thông tin và dữ liệu kết quả từ OCR sẽ được lưu trữ trong MySQL, nơi người dùng có thể tìm kiếm, chỉnh sửa hoặc tải về khi cần thiết.

Ưu điểm của Node.js và MySQL:

* Tốc độ xử lý cao: Node.js giúp ứng dụng có khả năng xử lý đồng thời nhiều yêu cầu mà không bị gián đoạn.
* Khả năng mở rộng linh hoạt: MySQL cho phép bạn mở rộng hệ thống theo yêu cầu mà không ảnh hưởng đến hiệu suất.
* Tính ổn định và bảo mật: Cả Node.js và MySQL đều là các công cụ ổn định và được sử dụng rộng rãi trong ngành công nghiệp, với các cộng đồng phát triển lớn và nhiều thư viện hỗ trợ.

Việc tích hợp Node.js với MySQL và Tesseract OCR sẽ giúp xây dựng một hệ thống web mạnh mẽ, có khả năng xử lý dữ liệu hình ảnh và lưu trữ kết quả số hóa một cách hiệu quả, hỗ trợ tốt cho các ứng dụng trong giáo dục và quản lý tài liệu.

# **CHƯƠNG II. PHÂN TÍCH THIẾT KẾ HỆ THỐNG**

## **2.1 Phân tích yêu cầu**

### 2.1.1 Mô tả nghiệp vụ

Website quản lý và upload dữ liệu bài giảng là một nền tảng hỗ hợ lưu trữ và quản lý tài liệu học tập. Mục tiêu của website là hỗ trợ giảng viên và sinh viên truy cập, tải lên và bảo mật tài liệu học tập hiệu quả. Về cơ bản nghiệp vụ bao gồm:

Quản lý tài liệu số hóa:

* Mục đích: Đáp ứng nhu cầu lưu trữ, quản lý và truy cập tài liệu học tập cho sinh viên và giảng viên.
* Quy trình nghiệp vụ:
  + Tải lên tài liệu từ giảng viên (bao gồm tệp PDF, slide, video).
  + Phân loại tài liệu theo môn học, chủ đề hoặc lớp học.
  + Tích hợp công cụ OCR (Tesseract OCR) để chuyển đổi tài liệu từ hình ảnh sang văn bản.

Quản lý quyền truy cập:

* Mục đích: Đảm bảo chỉ những người được phép (giảng viên, sinh viên) mới có thể truy cập tài liệu.
* Quy trình nghiệp vụ:
  + Đăng nhập, phân quyền sử dụng (quản trị viên, giảng viên, sinh viên).
  + Chỉ định quyền hạn như xem, chỉnh sửa, hoặc tải xuống tài liệu.

Bảo mật dữ liệu:

* Mục đích: Ngăn chặn rò rỉ thông tin và mất mát dữ liệu.
* Quy trình nghiệp vụ:
  + Mã hóa tài liệu trước khi lưu trữ trong cơ sở dữ liệu.
  + Triển khai sao lưu tự động và phục hồi dữ liệu khi cần thiết.
  + Ghi nhật ký hoạt động (audit logs) để theo dõi truy cập và hành động của người dùng.

Hỗ trợ học tập:

* Mục đích: Cung cấp nền tảng hỗ trợ học tập hiện đại, tiện lợi.
* Quy trình nghiệp vụ:
  + Sinh viên có thể truy cập tài liệu mọi lúc, mọi nơi.
  + Giảng viên có thể cập nhật hoặc thay đổi tài liệu dễ dàng

### 2.1.2 Yêu cầu chức năng

Chức năng cơ bản:

* Tải lên tài liệu:
  + Cho phép giảng viên tải lên các tài liệu giảng dạy.
  + Hỗ trợ nhiều định dạng tệp (PDF, DOC, video).
* Xử lý OCR:
  + Tích hợp Tesseract OCR để trích xuất văn bản từ hình ảnh.
* Tìm kiếm và phân loại tài liệu:
  + Sắp xếp tài liệu theo môn học hoặc lớp học.
  + Cung cấp thanh tìm kiếm để truy cập nhanh tài liệu cần thiết.

Chức năng bảo mật:

* Xác thực và phân quyền:
  + Đăng nhập bằng tài khoản sinh viên hoặc giảng viên.
  + Phân quyền truy cập dựa trên vai trò (giảng viên, sinh viên).
* Mã hóa tài liệu:
  + Áp dụng thuật toán mã hóa AES để bảo vệ tài liệu.
* Sao lưu dữ liệu:
  + Cung cấp chức năng sao lưu định kỳ tài liệu và cơ sở dữ liệu.

### 2.1.3 Yêu cầu phi chức năng

* Giao diện đơn giản, dễ sử dụng.
* Tối ưu hóa các chức năng cho người dùng.
* Hoạt động ổn định.
* Bảo mật tuyệt đối thông tin người dùng.

## **2.2 Phân tích hệ thống**

### 2.2.1 Xác định các tác nhân chính của hệ thống

* Người quản lý (Admin): Quản lý tài liệu, phân quyền truy cập và giám sát ghi lại các log hệ thống.
* Người dùng: Các chức năng về định danh tài khoản, tìm kiếm và các quyền với file được admin cung cấp.

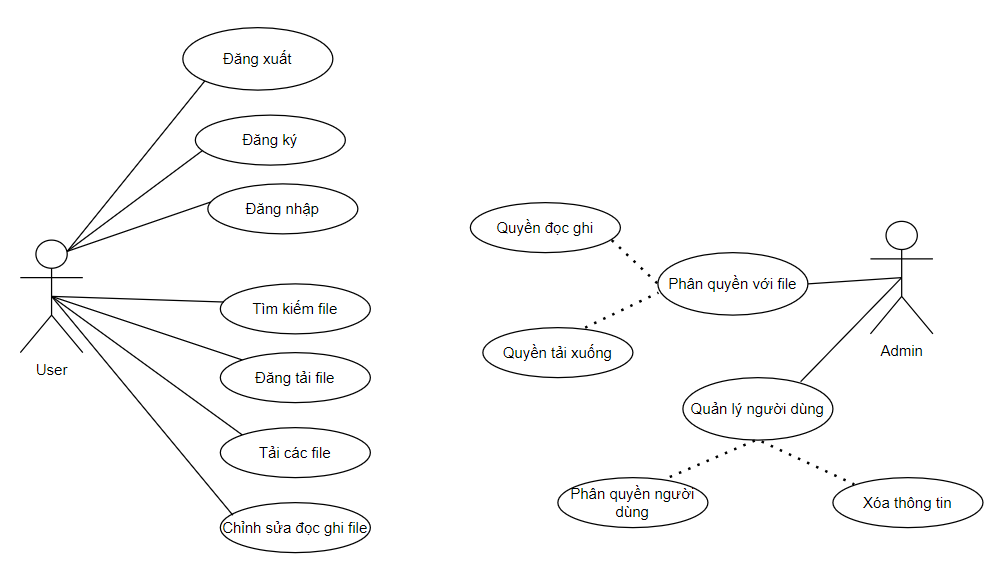
### 2.2.2 Các chức năng của tác nhân

|  |  |
| --- | --- |
| **Actor** | **Chức Năng** |
| Người dùng | Xem danh sách sản file: Khi người dùng truy cập vào website thì hệ thống sẽ tự động hiển thị các file đã đăng tải.  Tìm kiếm: Nếu tài liệu đó có tồn tại trong dữ liệu thì sẽ hiện thông tin tương ứng.  Đăng ký: Thực hiện chức năng đăng ký trên giao diện tương ứng.  Quyền đọc ghi với các thư mục: Theo quyền được admin cung cấp. |
| Người quản lý | Có tất cả chức năng của người dùng và thêm một số chức năng khác mà chỉ quyền admin mới thực hiện được:  Phân quyền người dùng  Phân quyền với các thư mục  Xem danh sách người dùng, thực hiện xóa nếu cần thiết |

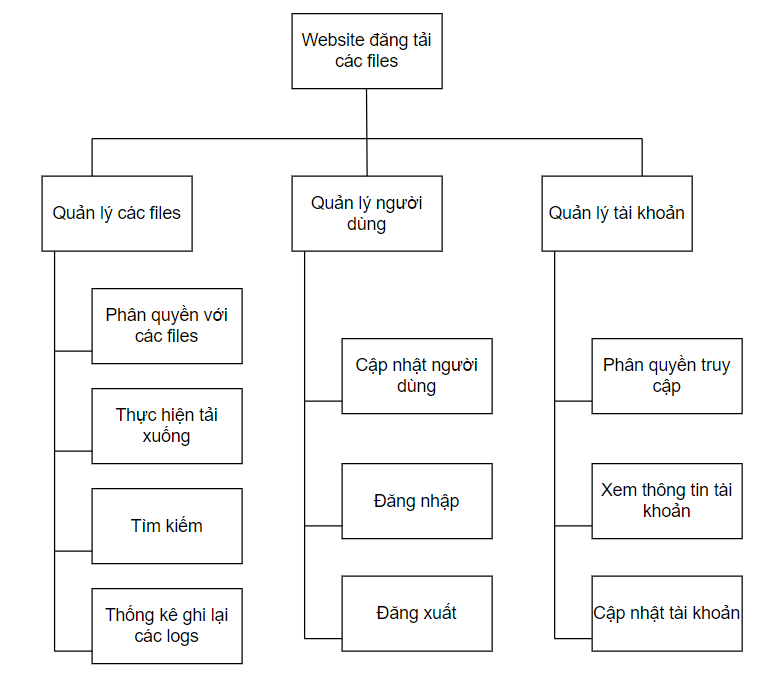
Bảng chi tiết danh sách các Use Case

|  |  |  |
| --- | --- | --- |
| **ID** | **Tên Usecase** | **Actor Sử Dụng** |
| 1 | Tìm kiếm, xem danh sách các tệp | Người dùng, người quản lý |
| 2 | Đăng ký, đăng nhập | Người dùng, người quản lý |
| 3 | Tìm kiếm | Người dùng, người quản lý |
| 5 | Phân quyền với các tệp | Người quản lý |
| 6 | Phân quyền với người dùng | Người quản lý |

### 2.2.3 Biểu đồ Use Case tổng quát



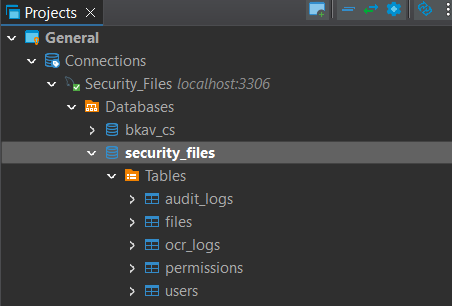
### 2.2.4 Biều đồ chức năng



## **2.3 Thiết kế hệ thống**

### 2.3.1 Thiết kế cơ sở dữ liệu

Thiết kế cơ sở dữ liệu Security\_Files bao gồm 5 bảng chứa các thông tin dữ liệu về: Người dùng, thông tin các tệp, thông tin các logs với tệp, thông tin các logs về phần mềm Tesseract OCR, và phân chia các quyền.



**Chi tiết các bảng dữ liệu và thuộc tính, các khóa chính, khóa ngoại, khóa duy nhất và mối quan hệ có trong cơ sở dữ liệu:**

**Bảng users:**

|  |  |  |  |
| --- | --- | --- | --- |
| Thuộc tính | Kiểu dữ liệu | Mô tả | Khóa |
| id | VARCHAR (50) | Mã định danh duy nhất cho người dùng | Khóa chính |
| fullname | VARCHAR (255) | Tên đầy đủ |  |
| username | VARCHAR (50) | Tên người dùng duy nhất | Khóa duy nhất |
| email | VARCHAR (100) | Địa chỉ email duy nhất | Khóa duy nhất |
| password | VARCHAR (255) | Mật khẩu được mã hóa |  |
| role | ENUM (‘user’, ‘admin’) | Quyền (mặc định là user) |  |
| created\_at | TIMESTAMP | Thời gian tạo |  |
| updated\_at | TIMESTAMP | Thời gian sửa đổi |  |

**Bảng files**

|  |  |  |  |
| --- | --- | --- | --- |
| Thuộc tính | Kiểu dữ liệu | Mô tả | Khóa |
| id | VARCHAR (255) | Mã định danh duy nhất cho file | Khóa chính |
| filename | VARCHAR (255) | Tên của file |  |
| filePath | VARCHAR (255) | Đường dẫn lưu trữ file |  |
| friendlyFileType | VARCHAR (255) | Tên thân thiện với file |  |
| formattedFileSize | VARCHAR (50) | Định dạng kích thước tệp cơ bản |  |
| uploadedAt | DATETIME | Thời gian tải file lên |  |
| ocrStatus | ENUM ‘'pending’, ‘processing’, ‘completed’, ‘failed’) | Trạng thái với OCR |  |
| ocrContent | TEXT | Nội dung trích xuất từ OCR |  |
| user\_id | VARCHAR (50) | Id người tải file lên | Khóa ngoại (users.id) |

**Bảng permissions**

|  |  |  |  |
| --- | --- | --- | --- |
| Thuộc tính | Kiểu dữ liệu | Mô tả | Khóa |
| id | VARCHAR (255) | Mã định danh duy nhất cho các quyền | Khóa chính |
| user\_id | VARCHAR (50) | Id người dùng | Khóa ngoại (users.id) |
| file\_id | VARCHAR (255) | Id của file | Khóa ngoại (files.id) |
| can\_view | BOOLEAN | Quyền được xem |  |
| can\_download | BOOLEAN | Quyền được tải xuống |  |
| can\_edit | BOOLEAN | Quyền được chỉnh sửa |  |
| created\_at | TIMESTAMP | Thời gian tạo |  |
| updated\_at | TIMESTAMP | Thời gian sửa đổi |  |

**Bảng audit\_logs**

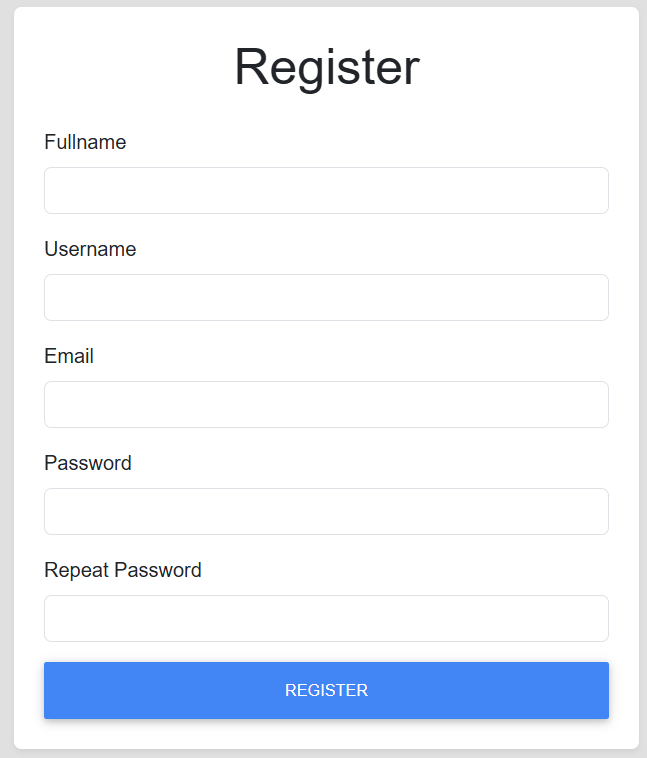
|  |  |  |  |
| --- | --- | --- | --- |
| Thuộc tính | Kiểu dữ liệu | Mô tả | Khóa |
| id | VARCHAR (255) | Mã định danh duy nhất cho các quyền | Khóa chính |
| user\_id | VARCHAR (50) | Id người dùng thực hiện hành động | Khóa ngoại (users.id) |
| file\_id | VARCHAR (255) | Id của file liên quan | Khóa ngoại (files.id) |
| action | ENUM ‘'upload’, ‘download’, ‘delete’, ‘edit’) | Hành động của người dùng |  |
| can\_download | BOOLEAN | Quyền được tải xuống |  |
| action\_timestamp | TIMESTAMP | Thời gian thực hiện |  |
| description | TEXT | Mô tả hành động |  |

**Bảng ocr\_logs**

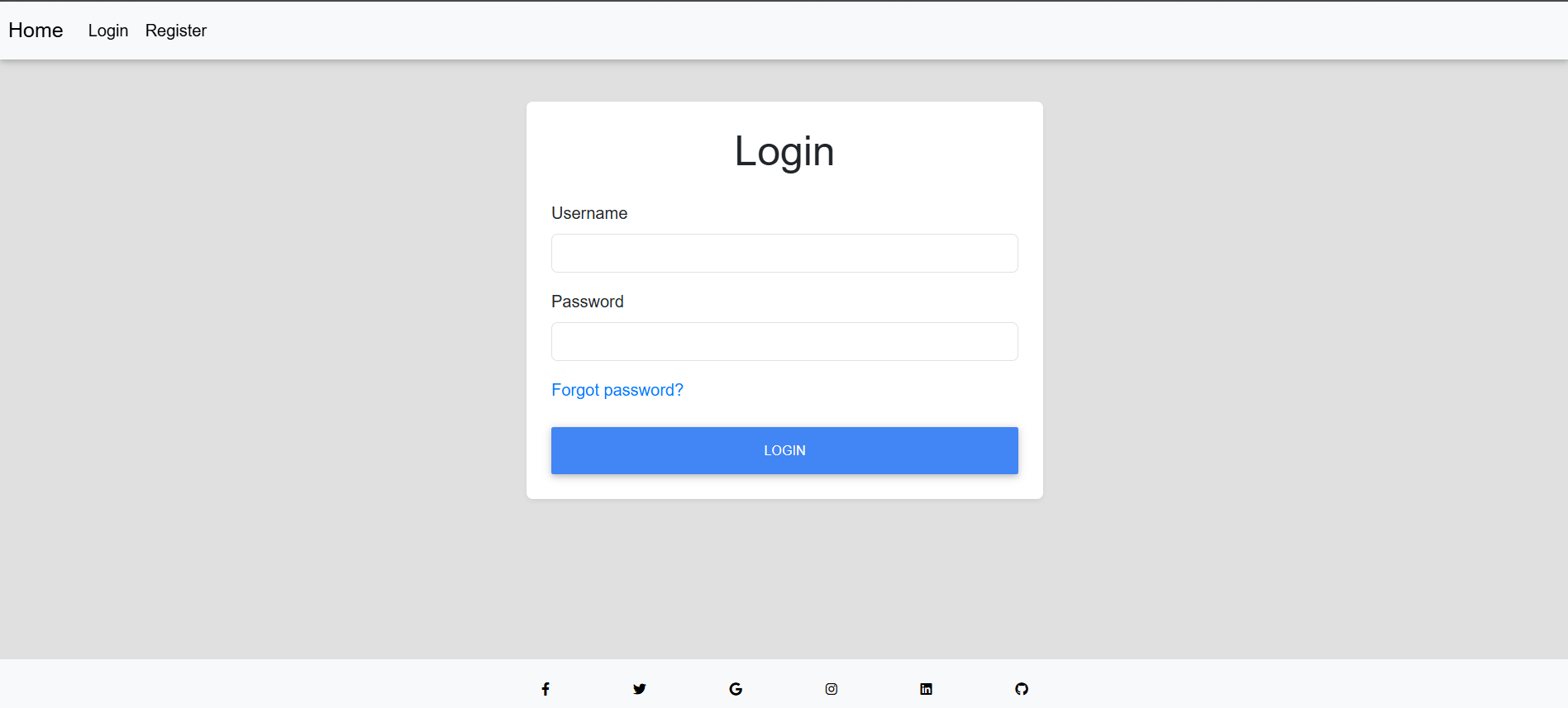
|  |  |  |  |
| --- | --- | --- | --- |
| Thuộc tính | Kiểu dữ liệu | Mô tả | Khóa |
| id | VARCHAR (255) | Mã định danh duy nhất cho OCR logs | Khóa chính |
| file\_id | VARCHAR (255) | Id của file | Khóa ngoại (files.id) |
| status | ENUM ‘'pending’, ‘processing’, ‘completed’, ‘failed’) | Trạng thái xử lý với OCR |  |
| result | TEXT | Kết quả xử lý OCR |  |
| processed\_at | TIMESTAMP | Thời gian xử lý OCR |  |
| error\_message | TEXT | Mô tả các lỗi của OCR |  |

### 2.3.2 Thiết kế các giao diện

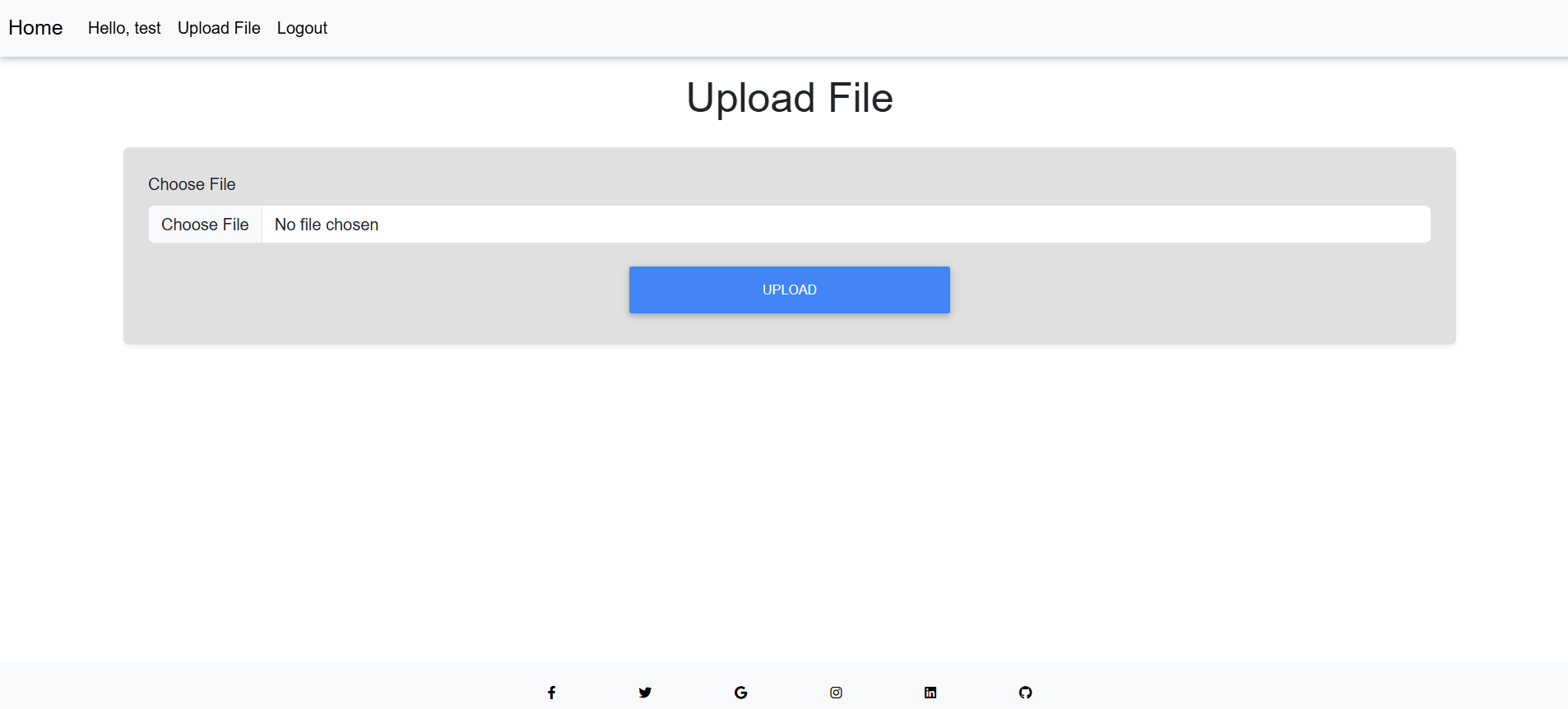
Giao diện đăng ký



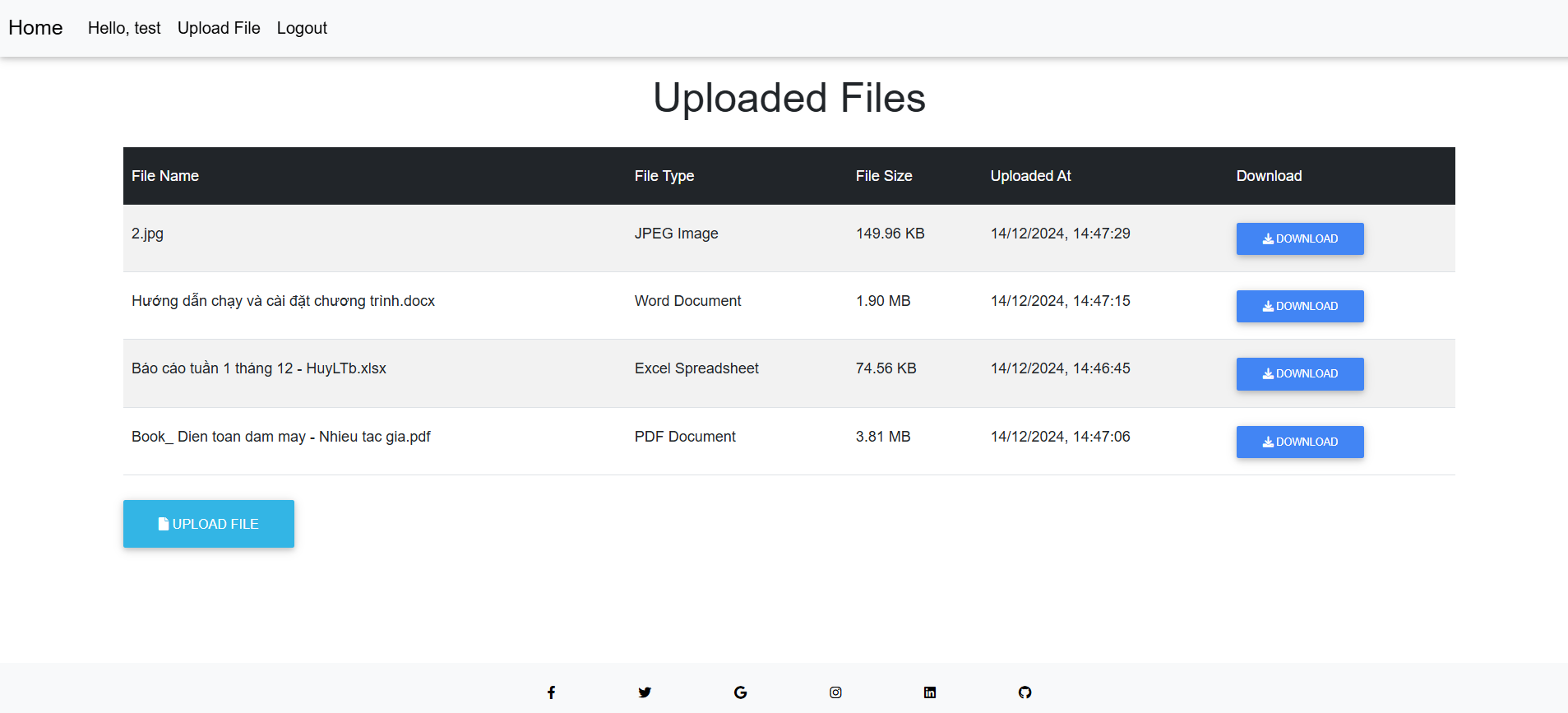
Giao diện đăng nhập



Giao diện tải các file



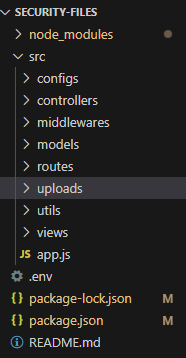
Giao diện hiển thị các file



# **CHƯƠNG III. XÂY DỰNG VÀ TRIỂN KHAI HỆ THỐNG**

## **3.1 Xây dựng các chức năng**

### 3.1.1 Cấu trúc chương trình

****

Chương trình được xây dựng theo mô hình MVC gồm các file chính mang tên của các tầng xử lý.

### 3.1.2 Các đối tượng và chức năng tương ứng

Các đối tượng:

**Người dùng (User):**

* Đăng ký tài khoản.
* Đăng nhập/Đăng xuất hệ thống.
* Upload và quản lý các file tài liệu.
* Xem danh sách file đã tải lên.
* Tải xuống các tệp.

**Người quản trị (Admin):**

* Giám sát, quản lý toàn bộ hệ thống và thông tin người dùng.
* Quản lý tài khoản người dùng (thêm, sửa xóa nếu cần).
* Quản lý tệp tin được upload
* Xóa tệp tin vi phạm.
* Xem danh sách tệp tin trên hệ thống.
* Theo dõi hoạt động hệ thống để đảm bảo an toàn và bảo mật.

Các chức năng theo các thư mục:

* **Thư mục configs:**
  + Chứa các file cấu hình hệ thống như kết nối cơ sở dữ liệu MySQL, thiết lập môi trường (dotenv), và các cài đặt chung.
* **Chức năng:**
  + Kết nối cơ sở dữ liệu MySQL.
  + Cấu hình cổng chạy server.
  + Thiết lập các biến môi trường từ file .env.
* **Thư mục controllers**
  + Chứa logic xử lý các yêu cầu từ người dùng và admin.
* **Chức năng:**
  + Xử lý yêu cầu đăng ký, đăng nhập.
  + Xử lý file upload (lưu trữ file vào thư mục uploads).
  + Trả về danh sách file đã upload cho người dùng.
  + Thực hiện các chức năng quản trị như xóa file hoặc quản lý tài khoản.
* **Thư mục middlewares:**
  + Chứa các middleware đảm bảo bảo mật và xác thực thông tin.
* **Chức năng:**
  + Xác thực JWT (JSON Web Token) cho phiên đăng nhập của người dùng.
  + Kiểm tra quyền truy cập cho người dùng và admin.
  + Thực hiện các biện pháp bảo mật chống tấn công như CORS hoặc CSRF.
* **Thư mục models:**
  + Chứa các định nghĩa bảng trong cơ sở dữ liệu MySQL sử dụng ORM hoặc truy vấn SQL.
* **Chức năng:**
  + Quản lý bảng người dùng (User): lưu trữ thông tin tài khoản.
  + Quản lý bảng file upload: lưu trữ thông tin file như tên, đường dẫn, và người tải.
* **Thư mục routes:**
  + Chứa các định nghĩa về endpoint API của hệ thống.
* **Chức năng:**
  + Định nghĩa các API về tài khoản, đăng tải, tìm kiếm và xử lý các tệp.
  + Phân tách rõ ràng giữa các route dành cho người dùng và admin.
* **Thư mục uploads:**
  + Chứa các file mà người dùng tải lên hệ thống.
* **Chức năng:**
  + Lưu trữ file với đường dẫn xác định.
  + Phân loại file theo user ID để dễ quản lý.
* **Thư mục utils:**
  + Chứa các hàm hỗ trợ như mã hóa mật khẩu, tạo JWT.
* **Chức năng:**
  + Mã hóa mật khẩu người dùng bằng **bcrypt**.
  + Tạo và xác minh JWT cho quá trình xác thực.
  + Hỗ trợ xử lý định dạng file tải lên.
* **Thư mục views:**
  + Chứa các file giao diện tương ứng.
* **Chức năng:**
  + Hiển thị giao diện trang đăng ký, đăng nhập.
  + Hiển thị danh sách file đã tải lên.
* **Thư mục app.js:**
  + File khởi chạy chính của ứng dụng.
* **Chức năng:**
  + Khởi động server Node.js.
  + Thiết lập route chính từ thư mục routes.
  + Kết nối tới MySQL thông qua file cấu hình trong configs.

## **3.2 Kiểm thử các chức năng**

### 3.2.1 Chức năng đăng nhập

Kịch bản kiểm thử: Người dùng nhập các thông tin đăng nhập, nếu thông tin sai thì đưa ra cảnh báo, còn thông tin đăng nhập chính xác sẽ vào bên trong.

Kết quả kiểm thử: Thành công, hệ thống cho phép người dùng truy cập các chức năng.

### 3.2.2 Chức năng tải tài liệu

Kịch bản kiểm thử: Người dùng sau khi chọn các file từ thiết bị cá nhân, tiến hành bấm nút đăng tải.

Kết quả kiểm thử: Thành công, hiển thị thông báo thành công, file tương ứng được tải lên và điều hướng tới trang lưu trữ các tệp hiện có.

### 3.2.3 Chức năng tìm kiếm

Kịch bản kiểm thử: Thực hiện các thao tác nhập dữ liệu trên thanh tìm kiếm.

Kết quả kiểm thử: Thành công, các thông tin về file được hiển thị, trường hợp không có file tương ứng sẽ hiển thị trang báo không có file cần tìm.

# **KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

## **Tổng kết đánh giá**

Thành quả đạt được:

* Về cơ bản đã xây dựng và hoàn thiện các chức năng cần có ở một trang web đăng tải các tệp tin.
* Tổ chức cấu trúc dự án trong thư mục rõ ràng, tách biệt các thành phần về giao diện, xử lý các logic và cơ sở dữ liệu.
* Tích hợp bảo mật, đảm bảo an toàn dữ liệu người dùng.
* Có thể dễ dàng quản lý việc thay đổi cấu trúc của website để phù hợp hơn với các xu hướng sau này.

Hạn chế:

* Giao diện còn chưa sinh động.
* Dữ liệu chưa được đầy đủ, chưa đúng với xu hướng thực tế.
* Khả năng xử lý các tệp tin định dạng không hợp lệ còn thiếu sót.
* Các chức năng còn chưa đầy đủ, một số chức năng cần tối ưu hơn nữa.

## **Hướng phát triển**

* Thiết kế trang web tiện dụng và có nhiều chức năng hơn.
* Tối ưu hóa được hiệu năng và tốc độ xử lý với các tệp lớn.
* Chỉnh sữa dữ liệu cho chính xác với thực tế và sẽ nâng cấp dần website để đáp ứng được những nhu cầu của khách hàng trong thời đại mới.

# **TÀI LIỆU THAM KHẢO**

[1] <https://en.wikipedia.org/>

[2] <https://stackoverflow.com/>

[3] <https://www.w3schools.com/>

[4] <https://nodejs.org/docs/latest/api/>

[5] <https://dev.mysql.com/doc/>