

TRƯỜNG ĐẠI HỌC ĐIỆN LỰC  
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO  
THỰC TẬP QUẢN TRỊ DỰ ÁN PHẦN MỀM

ĐỀ TÀI: XÂY DỰNG WEB NHẬN DIỆN VĂN BẢN TIẾNG  
VIỆT SỬ DỤNG TESSERACT

Sinh viên thực hiện	:: NGUYỄN VĂN NAM
Giảng viên hướng dẫn	:: HOÀNG THANH TÙNG
Mã sinh viên	:: 18810310428
Chuyên ngành	:: CÔNG NGHỆ PHẦN MỀM
Lớp	:: D13CNPM5
Khóa	:: 2018– 2023

Hà Nội, tháng 7 năm 2022

**ĐỀ CƯƠNG THỰC TẬP MÔN  
THỰC TẬP QUẢN TRỊ DỰ ÁN PHẦN MỀM**

**1. Tên đề tài:**

“Xây dựng web nhận diện văn bản tiếng việt sử dụng Tesseract”.

**2. Sinh viên thực hiện:**

Họ và tên: Nguyễn Văn Nam      MSSV: 18810310428  
Số điện thoại: 0398727881      Email: nguyenvannamtgdd35@gmail.com  
Vị trí thực tập: Lập trình viên .Net  
Thời gian thực tập: Từ ngày 04/04/2022 đến ngày 27/06/2022

**3. Giảng viên hướng dẫn :**

Họ và tên : Hoàng Thanh Tùng  
Số điện thoại : 0978421326      Email : tunght@epu.edu.vn

**4. Cán bộ hướng dẫn tại đơn vị thực tập**

Họ và tên: Lê Xuân Trường      Chức vụ: Quản lý dự án  
Số điện thoại : 0964038801      Email : truong.lexuan@vietis.com.vn  
Phòng/Bộ phận : Phòng sản xuất số 3  
Tên đơn vị thực tập : Công ty Cổ phần đầu tư và Giải pháp VietIS  
Địa chỉ: 3A Building, 82 Duy Tân, Hà Nội.

**4. Mô tả tóm tắt đề tài**

Xây dựng web nhận diện văn bản tiếng việt sử dụng công nghệ Tesseract, xây dựng giao diện cho web nhận diện.

**5. Nội dung hướng dẫn:**

- Chương 1: Giới thiệu công ty – đơn vị thực tập.
- Chương 2: Triển khai và thực hiện dự án.
- Chương 3: Kết thúc dự án

**6. Kết quả cần đạt được:**

- Phân chia công việc theo đầu mục.
- Xây dựng giao diện chương trình.

- Xây dựng được modul cho chương trình.

#### **7. Các yêu cầu đối với sinh viên:**

- Có khả năng học và tìm hiểu các tài liệu bằng Tiếng Anh.
- Có khả năng lập trình với Nodejs, ...
- Sử dụng công cụ lập trình Visual studio, Studio 3T, MongoDB, ...
- Sử dụng thành thạo trello.
- Có khả năng quản trị dự án.

**Giảng viên hướng dẫn**

**Hoàng Thanh Tùng**

**Sinh viên thực hiện**

**Nguyễn Văn Nam**

## PHIẾU ĐIỂM

**Sinh viên thực hiện:**

Họ và tên	Nhiệm vụ	Chữ ký
Nguyễn Văn Nam 18810310428		

**Giảng viên chấm:**

Họ và tên	Nhận xét	Điểm
Giảng viên chấm 1:		
Giảng viên chấm 2:		

## MỤC LỤC

CHƯƠNG 1: GIỚI THIỆU CÔNG TY - ĐƠN VỊ THỰC TẬP.....	1
1.1. Giới thiệu về VIETIS .....	1
1.2. Lịch sử hình thành .....	3
1.3. Tầm nhìn, sứ mệnh, giá trị cốt lõi .....	3
1.3.1. Tầm nhìn – sứ mệnh – giá trị cốt lõi.....	3
1.3.2. Văn hóa doanh nghiệp .....	3
CHƯƠNG 2: TRIỂN KHAI VÀ THỰC HIỆN DỰ ÁN.....	6
2.1. Đề cương dự án .....	6
2.1.1. Tính cập thiết của dự án.....	6
2.1.2. Mục tiêu của đề cương.....	7
2.1.3. Nội dung .....	7
2.1.4. Hiệu quả dự kiến .....	8
2.1.5. Dự kiến tiến trình triển khai nhiệm vụ.....	8
2.1.6. Kết luận.....	9
2.2. Hồ sơ dự án khả thi.....	9
2.2.1. Khảo sát hiện trạng .....	9
2.2.2. Đánh giá các giải pháp về mặt kỹ thuật tài chính .....	10
2.2.3. Kế hoạch tổ chức, thực hiện nhiệm vụ .....	11
2.2.4. Dự toán và danh sách rủi ro .....	11
2.2.4.1. Dự toán tổng quan.....	11
2.2.4.2. Dự toán chi tiết.....	12
2.2.6. Kế hoạch dự án .....	15
2.2.6.1. Phân công nhiệm vụ .....	15

2.2.6.2. Kế hoạch dự án.....	15
2.2.7. Phân tích hệ thống.....	16
2.2.8. Rủi ro dự án .....	18
2.2.8.1. Rủi ro trong công nghệ.....	18
2.2.8.2. Rủi ro do con người .....	19
2.2.8.3. Rủi ro trong quá trình nghiệp vụ.....	20
2.2.8.4. Chi tiết các loại rủi ro .....	21
2.3. Công cụ quản lý dự án.....	26
2.3.1. Công cụ trello.....	26
2.3.2. Áp dụng trong quản lý dự án.....	27
2.4. Mô hình quản lý dự án.....	27
2.5. Tự nhận xét và đánh giá .....	27
<b>CHƯƠNG 3: KẾT THÚC DỰ ÁN .....</b>	<b>29</b>
3.1. Tài liệu kết thúc dự án .....	29
3.1.1. Mô tả ngắn về dự án .....	29
3.1.2. Lý do phát triển dự án.....	29
3.1.3. Kết quả đạt được.....	29
3.1.4. Đánh giá về kết quả dự án .....	29
3.1.5. Một số vấn đề khi kết thúc dự án.....	29
3.2. Kết thúc dự án .....	30
3.3. Tài liệu đào tạo .....	32
<b>KẾT LUẬN .....</b>	<b>33</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>34</b>

## **DANH MỤC HÌNH ẢNH**

Hình 1.1: Công ty Cổ phần Đầu tư và Giải pháp VietIS .....	1
Hình 1.2: Các đối tác tiêu biểu của VietIS Corporation .....	1
Hình 1.3: Seminar nghề BrSE.....	2
Hình 1.4: Sự kiện Teambuilding.....	4
Hình 1.5: VietIS tổ chức chương trình trải nghiệm tiếng Nhật cho nhân viên.....	4
Hình 1.6: Hình ảnh cho biết doanh nghiệp có nhiều cơ hội làm việc thực tập hấp dẫn đang chờ đợi học viên FUNiX.....	5
Hình 2.1: Áp dụng trello trong quản lý dự án.....	27

## **DANH MỤC BẢNG BIỂU**

Bảng 2.1: Dự toán tổng quan dự án .....	11
Bảng 2.2: Dự toán chi tiết dự án .....	13
Bảng 2.3: Phân công nhiệm vụ .....	15
Bảng 2.4: Kế hoạch dự án.....	15
Bảng 2.5: Rủi ro trong công nghệ.....	18
Bảng 2.6: Rủi ro do con người.....	19
Bảng 2.7: Rủi ro trong quá trình nghiệp vụ .....	20
Bảng 2.8: Nhận xét và đánh giá công việc .....	28

## LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với các thầy cô của trường **Đại học Điện Lực**, đặc biệt là các thầy cô **Công Nghệ Thông Tin** của trường đã tạo điều kiện cho em thực hiện báo cáo. Và em cũng xin chân thành cảm ơn thầy **Hoàng Thanh Tùng** đã nhiệt tình hướng dẫn em hoàn thành tốt báo cáo.

Trong quá trình thực tập, cũng như là trong quá trình làm bài báo cáo thực tập, khó tránh khỏi sai sót, rất mong các thầy, cô bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, em rất mong nhận được ý kiến đóng góp thầy, cô để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn bài báo cáo.

Em xin chân thành cảm ơn!

## LỜI NÓI ĐẦU

Trong vài thập kỉ trở lại đây, với sự bùng nổ của ngành công nghệ thông tin đã mang lại cho chúng ta những thành tựu công nghệ mới, việc này tạo điều kiện cho sự phát triển và ra đời của công nghệ AI trong ứng dụng văn phòng cũng như tòa nhà, việc quản lý nhân sự...

Dựa trên nền tảng đó, em đã kết hợp sử dụng công nghệ nhận diện văn bản để “**Xây dựng website nhận dạng văn bản tiếng việt sử dụng Tesseract**”.

Trong quá trình này, với sự giúp đỡ và tư vấn nhiệt tình từ thầy **Hoàng Thanh Tùng**, cùng với các thầy cô và toàn thể các anh chị đồng nghiệp tại **Công ty Cổ phần đầu tư và Giải pháp VietIS** đã giúp em xây dựng sản phẩm này.

Em xin chân thành cảm ơn!

## CHƯƠNG 1: GIỚI THIỆU CÔNG TY - ĐƠN VỊ THỰC TẬP

### 1.1. Giới thiệu về VIETIS



Hình 1.1: Công ty Cổ phần Đầu tư và Giải pháp VietIS

Công ty Cổ Phàn Đầu tư và Giải pháp VietIS (gọi tắt là VietIS Corporation) được thành lập từ năm 2009 và bắt đầu cung cấp dịch vụ phát triển phần mềm do đối tác Nhật Bản từ năm 2013.

VietIS Corporation hiện có 1 trung tâm phát triển tại Hà Nội và 1 văn phòng tại Tokyo Nhật Bản với đội ngũ kỹ sư lên đến hơn 250 người. VietIS luôn theo đuổi mục tiêu tạo ra những đột phá và trở thành một trong những công ty công nghệ hàng đầu Việt Nam, là đối tác đáng tin cậy của khách hàng Nhật Bản.

Sở hữu đội ngũ kỹ sư, chuyên gia và quản lý dự án nhiều năm kinh nghiệm tham gia phát triển ứng dụng điện thoại thông minh, dịch vụ ứng dụng, phát triển hệ thống cốt lõi và phát triển công nghệ tiên tiến như IoT, AI và điện toán đám mây.

VietIS luôn là đối tác làm việc của nhiều khách hàng Nhật Bản lớn như Honda, NEC, Fujitus, JBS, Hoya... Với định hướng tăng trưởng mạnh về số lượng nhân sự ở cả Việt Nam và Nhật Bản, VietIS hiện có nhiều cơ hội việc làm, thực tập hấp dẫn chờ đợi học viên FUNiX.

#### ĐỐI TÁC TIÊU BIỂU



Hình 1.2: Các đối tác tiêu biểu của VietIS Corporation

Trong bối cảnh Covid – 19, công ty vẫn liên tục tuyển mới các vị trí từ nhân viên đến quản lý với mức lương tới 35 triệu đồng một tháng. Để đối phó với dịch bệnh, công ty tổ chức phỏng vấn online qua skype từ ngày 09/03, làm việc từ xa từ ngày 01/04. Những vị trí mà công ty liên tục tuyển dụng gồm: Junior Java, Mobile Techlead, BrSE Mobile, Junior Tester, Sale IT, ...

Cùng với mức lương hấp dẫn, công ty có nhiều chương trình đào tạo giúp ứng viên phát triển như: Đào tạo Fresher cho sinh viên mới ra trường: đào tạo quy trình sản xuất phần mềm – Software Process Development cho nhân viên mới ra nhập; các khóa đào tạo Project Management, Khóa đào tạo kỹ năng mềm: Horensco, Leadership, ... Khóa đào tạo ngôn ngữ: Tiếng Nhật (N5), tiếng anh, ...

Ngoài ra, VietIS thường xuyên tổ chức seminar chia sẻ công nghệ giúp cán bộ, nhân viên nâng cao tri thức, trình độ. Cùng với đó công ty có chính sách hỗ trợ nhân viên học và thi các chứng chỉ cần thiết trong công việc cùng với mức trợ cấp hấp dẫn.



Hình 1.3: Seminar nghề BrSE

## **1.2. Lịch sử hình thành**

Công ty Cổ phần Đầu tư và Giải pháp VietIS có 2 chi nhánh hiện nay, đó là một chi nhánh tại Hà Nội và một chi nhánh tại Nhật Bản.

Chi Nhánh tại Nhật Bản có địa chỉ là Tokyo, Shinagawa, Nishigotanda, 2-25-1 Intex 2F. Gồm các kỹ sư hàng đầu tại VietIS.

Với Chi nhánh tại Hà Nội, hiện nay VietIS có 2 địa điểm chính là tầng 3 và tầng 5, tòa nhà 3A, ngõ 82 Duy Tân và tầng 7 Cầu Giấy, Hà Nội.

## **1.3. Tầm nhìn, sứ mệnh, giá trị cốt lõi**

### **1.3.1. Tầm nhìn – sứ mệnh – giá trị cốt lõi**

VietIS Corporation luôn mong muốn trở thành một trong những công ty công nghệ hàng đầu Việt Nam, là đối tác đáng tin cậy đối với khách hàng Nhật Bản.

VietIS là một công ty Outsourcing chuyên cung cấp các giải pháp phần mềm như Cloud Service, Business Application Development, Web/Mobile Development, ... cho thị trường Nhật Bản. Chính vì vậy, VietIS luôn mong muốn trở thành một đối tác đáng tin cậy với tất cả các đối tác.

### **1.3.2. Văn hóa doanh nghiệp**

**Môi trường và các hoạt động:** VietIS hiểu rằng môi trường làm nên con người, cho nên VietIS luôn chú trọng để phát triển môi trường học hỏi, vui chơi tốt nhất cho nhân viên.

Với môi trường làm việc trẻ trung, năng động và chuyên nghiệp và nhiều cơ hội thăng tiến. Sức trẻ và nhiệt huyết của VietIS được tạo nên từ chính mỗi cá nhân công ty, cho nên VietIS không ngừng thay đổi để hoàn thiện hơn nữa bản thân.

Với việc tổ chức sinh nhật tháng của nhân viên, cũng như các hoạt động khác do công đoàn tổ chức như Teambuilding ngoài trời định kì, các hoạt động vui chơi giải trí khác, tham gia vào các câu lạc bộ vui chơi như CLB bóng đá, ... đã tạo ra một môi trường năng động phù hợp với rất nhiều bạn trẻ.



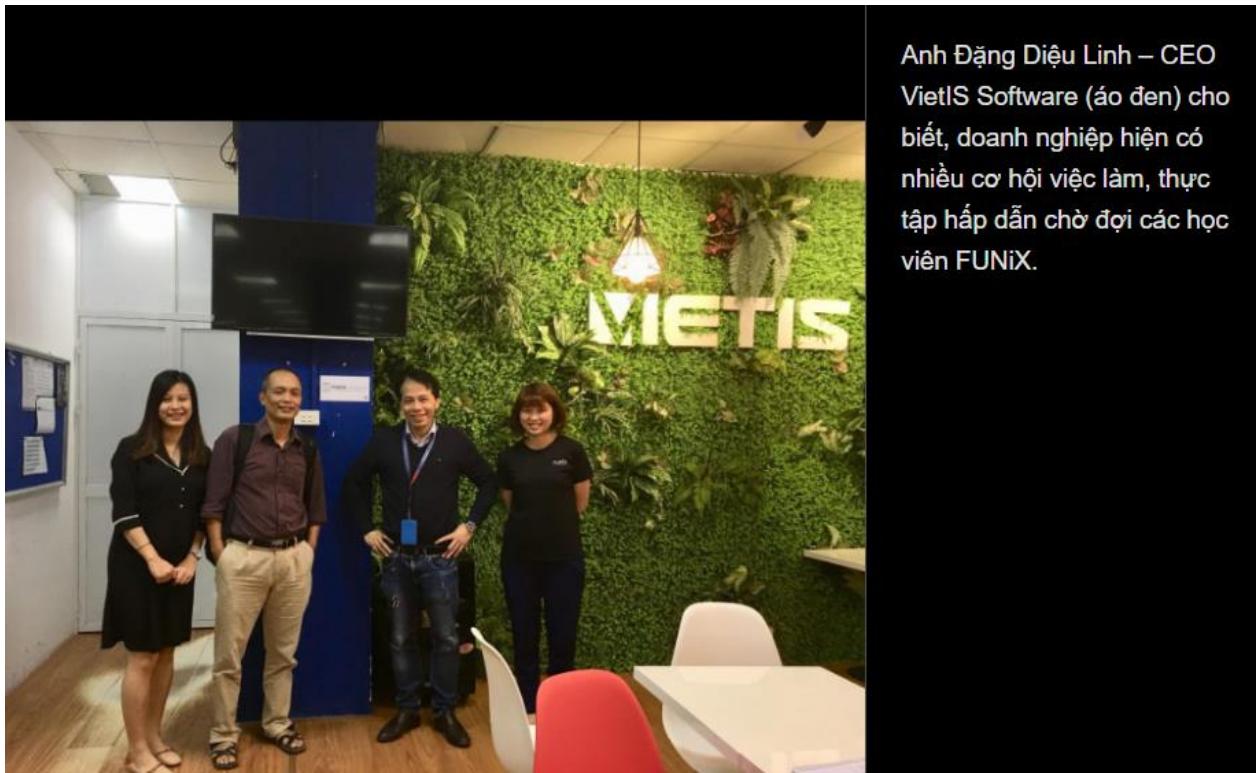
Hình 1.4: Sự kiện Teambuilding



Hình 1.5: VietIS tổ chức chương trình trải nghiệm tiếng Nhật cho nhân viên

**VietIS dành chỗ cho sinh viên tới thực tập:** Công ty VietIS còn thiết kế riêng một góc với 20 chỗ ngồi cho học viên FUNiX tới thực tập. VietIS là đơn vị ký kết hợp tác đào tạo – tuyển dụng với FUNiX trong tháng ba, thuộc chương trình “100 doanh nghiệp IT cùng FUNiX phát triển nguồn nhân lực”

Hợp tác nhằm đảm bảo bám sát và đáp ứng nhu cầu tuyển dụng thực tế của doanh nghiệp, phù hợp xu hướng phát triển năng lực lao động của ngành phần mềm toàn cầu. VietIS sẵn sàng cử chuyên gia có trình độ cao tham gia trực tiếp vào đào tạo trong vai trò cố vấn chuyên môn học tập (mentor) để đồng hành cùng sinh viên trong suốt quá trình học tập.



Anh Đặng Diệu Linh – CEO  
VietIS Software (áo đen) cho  
biết, doanh nghiệp hiện có  
nhiều cơ hội việc làm, thực  
tập hấp dẫn chờ đợi các học  
viên FUNiX.

Hình 1.6: Hình ảnh cho biết doanh nghiệp có nhiều cơ hội làm việc  
thực tập hấp dẫn đang chờ đợi học viên FUNiX

## CHƯƠNG 2: TRIỂN KHAI VÀ THỰC HIỆN DỰ ÁN

### 2.1. Đề cương dự án

**Tên dự án:** “Xây dựng Web nhận dạng văn bản tiếng việt sử dụng Tesseract”.

#### 2.1.1. Tính cấp thiết của dự án

Nhận dạng văn bản tiếng việt là một yêu cầu cấp thiết đối với Việt Nam. Hiện nay các loại tài liệu tăng với số lượng khổng lồ, và việc lưu trữ được các loại tài liệu là một vấn đề cấp thiết đang được đặt ra cho các công ty, các doanh nghiệp, ... Chính điều này đang làm cho việc áp dụng bài toán nhận dạng văn bản tiếng việt trở nên khó khăn.

Khi gặp một vấn đề mới, chúng ta thường giải quyết nó dựa vào những tri thức, kinh nghiệm có trước. Ví dụ như giải một bài toán ta thường liên hệ để đưa chúng về các dạng bài toán đã làm trước đây hoặc tìm sự tương đồng trong các bài toán. Việc áp dụng tri thức này làm tăng tốc cũng như chất lượng của việc học. Nhận xét này không chỉ liên quan đến việc học của con người mà còn liên quan đến học máy. Việc học trong một nhiệm vụ mới được cải thiện bằng việc sử dụng tri thức đã được lưu tại từ những nhiệm vụ học trước đó. Nói cách khác là ta sử dụng những tri thức đã có nhằm nâng cao hiệu quả của việc học cho nhiệm vụ mới.

Hiện nay việc rút trích từ ngữ từ hình ảnh đang ngày càng phát triển, bên cạnh sự gia tăng về nhu cầu là sự phát triển nhận dạng ký tự quang học hay còn được gọi tắt là OCR. Đây là một công nghệ giúp chuyển đổi hình ảnh của chữ viết tay hay chữ đánh máy thành các ký tự đã được mã hóa trong máy tính.

Giả sử chúng ta cần chỉnh sửa một số tài liệu giấy như: các bài viết trên tạp chí, tờ rơi hoặc một tập tin PDF hình ảnh. Rõ ràng chúng ta không thể sử dụng máy quét để chuyển các tài liệu này thành tập tin văn bản để có thể chỉnh sửa. Tất cả những gì máy quét có thể làm là tạo ra một hình ảnh hoặc một bản chụp của các tài liệu. Để giải nén và sử dụng lại dữ liệu từ tài liệu được quét, hình ảnh máy ảnh và hình ảnh của các tập tin PDF chúng ta cần một phần mềm OCR. Nó sẽ xuất ra ký tự trên hình ảnh, ghép chúng thành các từ và sau đó ghép các từ thành các câu. Nhờ vậy, chúng ta có thể truy cập và chỉnh sửa nội dung của tài liệu gốc.

Tương tự, những tài liệu cũ đang bị hư hại theo thời gian và việc viết tay hay đánh máy lại những tài liệu đó sẽ tốn rất nhiều thời gian cũng như chi phí cho việc này, ngoài ra độ chính xác còn không được đảm bảo cũng như sự an toàn cho tài liệu nền. Việc này rất cần một công nghệ lấy từ ngữ từ hình ảnh chụp.

Trong bài báo cáo này, tôi sẽ tìm hiểu, chỉnh sửa công cụ Tesseract để thực thi được việc rút trích từ các văn bản từ tập tin hình ảnh cũng như từ các tập tin PDF.

### **2.1.2. Mục tiêu của đề cương**

Mục tiêu chính của đề cương là xây dựng web nhận dạng văn bản tiếng Việt sử dụng thuật toán Tesseract.

Hỗ trợ khắc phục những khó khăn gặp phải do đặc điểm của văn bản tiếng Việt, đồng thời rút ngắn thời gian cũng như lưu trữ được các loại tài liệu lâu hơn và bảo mật hơn.

**Phạm vi đề cương:** có thể áp dụng cho các doanh nghiệp, công ty, trường học, cửa hàng, ...

### **2.1.3. Nội dung**

Xây dựng web nhận diện văn bản tiếng Việt sử dụng công nghệ Tesseract với:

Các tính năng:

- Sử dụng công nghệ Tesseract được dùng phổ biến trong các kỹ thuật OCR, nhằm cải tiến độ chính xác nhận diện.
- Dễ dàng sử dụng cho mọi đối tượng như các doanh nghiệp, công ty nhỏ và vừa, các trường học, trung cư, ...
- Giao diện thân thiện với người dùng.
- Tối ưu hóa việc cải thiện lưu trữ các loại tài liệu cũ, có nguy cơ bị nguy hại đồng thời cũng cho phép việc chỉnh sửa dữ liệu thông tin của các tập tin PDF, các tập tin hình ảnh trên dữ liệu nền.

Chức năng của hệ thống:

- Đăng nhập, đăng xuất tài khoản có sử dụng phân quyền giúp cho người dùng cũng như người quản trị có các quyền riêng của mình.

- Quản lý đưa ra các loại chuyển đổi: tức là các loại convert ví dụ như từ hình ảnh với đầu vào là các file ảnh như file .jpg, .png, ... chuyển đổi đầu ra sẽ là một file txt hay chuyển đổi từ dạng file .pdf sang file txt. Ngoài ra quản trị còn có chức năng thêm, sửa, xóa các loại chuyển đổi này.
- Quản lý hóa đơn: cung cấp cho admin việc quản lý hóa đơn bao gồm các chức năng như thêm, sửa, xóa thông tin hóa đơn khi khách hàng đăng ký dịch vụ.
- Quản lý thông tin các file người dùng: Kiểm soát được số lượng file mà người dùng đã sử dụng, ...
- Ngoài ra, còn có một số chức năng khác.

Ưu điểm và nhược điểm:

- Dễ dàng chuyển đổi dữ liệu từ file ảnh hay file pdf sang file txt mà không mất nhiều thời gian, công sức cũng như đảm bảo về độ chính xác.
- Theo dõi, kiểm soát các file mà người dùng sử dụng, ...
- Cải tiến về độ chính xác.

#### **2.1.4. Hiệu quả dự kiến**

- Tính sẵn sàng của hệ thống, dễ dàng sử dụng.
- Vận hành ổn định, nhanh chóng.
- Truy xuất dữ liệu, chuyển đổi dữ liệu giữa các dạng chuyển đổi nhanh chóng, chính xác.
- Quản lý thông tin một cách đầy đủ, chính xác và chi tiết.
- Đánh giá thống kê, báo cáo, review chính xác.

#### **2.1.5. Dự kiến tiến trình triển khai nhiệm vụ**

- Tìm hiểu về công nghệ Tesseract.
- Tham khảo các công nghệ, kỹ thuật OCR khác.
- Xây dựng chương trình.
- Áp dụng và triển khai.

## **2.1.6. Kết luận**

- Có thể thay đổi phương pháp sửa dữ liệu nền, ... hiện tại bằng phương pháp nhận diện văn bản tiếng việt sử dụng công nghệ Tesseract.
- Có thể áp dụng trong giai đoạn thử nghiệm.
- Nếu được thực hiện sẽ tối ưu thời gian chuyển đổi các loại định dạng file từ ảnh, pdf sang định dạng file txt một cách nhanh chóng, tối đa.

## **2.2. Hồ sơ dự án khả thi**

### **2.2.1. Khảo sát hiện trạng**

#### **Hiện trạng hiện nay:**

Hiện nay việc rút trích từ ngữ từ hình ảnh đang ngày càng phát triển, bên cạnh sự gia tăng về nhu cầu là sự phát triển nhận dạng ký tự quang học hay còn được gọi tắt là OCR. Đây là một công nghệ giúp chuyển đổi hình ảnh của chữ viết tay hay chữ đánh máy thành các ký tự đã được mã hóa trong máy tính.

Giả sử chúng ta cần chỉnh sửa một số tài liệu giấy như: các bài viết trên tạp chí, tờ rơi hoặc một tập tin PDF hình ảnh. Rõ ràng chúng ta không thể sử dụng máy quét để chuyển các tài liệu này thành tập tin văn bản để có thể chỉnh sửa. Tất cả những gì máy quét có thể làm là tạo ra một hình ảnh hoặc một bản chụp của các tài liệu. Để giải nén và sử dụng lại dữ liệu từ tài liệu được quét, hình ảnh máy ảnh và hình ảnh của các tập tin PDF chúng ta cần một phần mềm OCR. Nó sẽ xuất ra kí tự trên hình ảnh, ghép chúng thành các từ và sau đó ghép các từ thành các câu. Nhờ vậy, chúng ta có thể truy cập và chỉnh sửa nội dung của tài liệu gốc.

Tương tự, những tài liệu cũ đang bị hư hại theo thời gian và việc viết tay hay đánh máy lại những tài liệu đó sẽ tốn rất nhiều thời gian cũng như chi phí cho việc này, ngoài ra độ chính xác còn không được đảm bảo cũng như sự an toàn cho tài liệu nền. Việc này cần một công nghệ lấy từ ngữ từ hình ảnh chụp.

Trong khuôn khổ đề tài báo cáo thực tập này, mục tiêu chính là để rèn luyện, thực hiện xây dựng hệ thống với những nhiệm vụ cơ bản.

## **Ưu điểm và nhược điểm:**

- **Ưu điểm:**
  - Người dùng đã quen với việc khi mà muốn xuất dữ liệu từ file ảnh thì tự gõ, đánh máy, ...
  - Phương pháp cũng tương đối ổn định đối với người dùng.
- **Nhược điểm:**
  - Người dùng có khi sẽ bị lỗi dữ liệu khi mà chuyển từ dữ liệu dạng ảnh sang dữ liệu dạng text thông qua việc đánh máy.
  - Mất thời gian, mà có khi dữ liệu nền bị ảnh hưởng.

## **Đề xuất giải pháp:**

- Thay thế việc chuyển đổi dữ liệu từ dạng ảnh, dạng pdf sang dạng text từ việc sử dụng gõ tay, ... với nhiều nhược điểm, ảnh hưởng đến độ chính xác thông tin của dữ liệu sang việc sử dụng web nhận diện đơn giản, dễ sử dụng, tính chính xác và tốc độ xử lý nhanh.

## **Yêu cầu hệ thống dự kiến:**

- Thực hiện tốt các chức năng hiện hành.
- Tuân thủ việc chuyển đổi các loại dữ liệu.
- Tốc độ xử lý nhanh và có độ chính xác cao theo thời gian thực.
- Linh hoạt trong việc sử dụng.
- Giao diện thân thiện, dễ sử dụng.
- Bảo mật tốt, chỉ làm việc với người có quyền sử dụng.

## **Yêu cầu về môi trường:**

- Hệ điều hành: window
- Hệ quản trị cơ sở dữ liệu: MongoDB.

### **2.2.2. Đánh giá các giải pháp về mặt kỹ thuật tài chính**

- Về mặt kỹ thuật: giúp quản lý một cách dễ dàng, nhanh chóng, chính xác từ khâu chuyển đổi dữ liệu từ dạng dữ liệu này sang dạng dữ liệu kia, giúp tiết kiệm thời gian, hạn chế việc sai sót về dữ liệu sau khi chuyển đổi.

- Về mặt tài chính: giúp cho công ty tiết kiệm được thời gian, công sức đào tạo nhân viên về việc xử lý, lưu trữ, chỉnh sửa tài liệu tránh nhầm lẫn khi quản lý dữ liệu.

### **2.2.3. Kế hoạch tổ chức, thực hiện nhiệm vụ**

- Dự kiến phân công trách nhiệm giữa các đơn vị:
  - Chia thành các đơn vị nghiệp vụ, phân tích, lập trình, kiểm thử và quản lý.
  - Các đơn vị có các trách nhiệm và nhiệm vụ riêng, vì vậy tất cả phải tự giác hoàn thành công việc, bên cạnh đó cần phân phối và giúp đỡ nhau để hoàn thành nhiệm vụ tốt nhất.
- Dự kiến thành lập đội dự án để dễ quản lý và triển khai.

### **2.2.4. Dự toán và danh sách rủi ro**

#### **2.2.4.1. Dự toán tổng quan**

STT	Nội dung công việc	Chi Phí	Ghi chú
1	Khởi động DA	5.000.000	
2	Khảo sát thực tế	5.000.000	
3	Phân tích và xác định yêu cầu hệ thống	9.000.000	
4	Thiết kế CSDL	8.000.000	
5	Thiết kế giao diện	3.500.000	
6	Lập trình	32.000.000	
7	Kiểm thử	15.000.000	
8	Kết thúc dự án	5.000.000	
9	Chi phí dự phòng	20.000.000	
12	Chi phí cho hội họp, tiếp khách	3.000.000	
13	Chi phí cho các vấn đề công nghệ kỹ thuật thực hiện dự án	20.000.000	
<b>Tổng</b>		<b>125.500.000</b>	

Bảng 2.1: Dự toán tổng quan dự án

#### 2.2.4.2. Dự toán chi tiết

STT	Chi tiết	Đơn giá (VNĐ)	Thời gian (Giờ)	Thành tiền (VNĐ)
1	<b>Khảo sát</b>	<b>150.000</b>	<b>20</b>	<b>5.000.000</b>
	- Quan sát			1.000.000
	- Phòng vấn			1.500.000
	- Điều tra thăm dò			2.500.000
2	<b>Phân tích chức năng</b>	<b>150.000</b>	<b>60</b>	<b>9.000.000</b>
	- Đăng nhập, đăng xuất			1.000.000
	- Chuyển đổi các loại dữ liệu			3.500.000
	- Quản lý hóa đơn			3.000.000
	- Quản lý file người dùng			1.500.000
3	<b>Thiết kế cơ sở dữ liệu</b>	<b>150.000</b>	<b>20</b>	<b>8.000.000</b>
	- collection users			1.000.000
	- collection reports			500.000
	- collection reviews			500.000
	- collection bills			1.500.000
	- collection accounts			1.000.000
	- collection converts			2.500.000
	- collection images			1.000.000
4	<b>Thiết kế giao diện</b>	<b>150.000</b>	<b>30</b>	<b>3.500.000</b>
	- Giao diện chương trình			3.500.000
5	<b>Lập trình và tích hợp hệ thống</b>	<b>150.000</b>	<b>80</b>	<b>32.000.000</b>
	- Xây dựng các bảng CSDL			5.000.000

	- Xây dựng chức năng quản trị người dùng			3.000.000
	- Lập trình module quản trị hệ thống			7.000.000
	- Lập trình chức năng quản lý			9.500.000
	- Lập trình module thực hiện project			3.000.000
	- Lập trình module thực hiện kế hoạch			2.000.000
	- Tích hợp hệ thống			2.000.000
<b>6</b>	<b>Kiểm thử và sửa lỗi</b>	<b>150.000</b>	<b>20</b>	<b>15.000.000</b>
	- Kiểm thử mức đơn vị			3.000.000
	- Kiểm thử mức tích hợp			3.000.000
	- Kiểm thử mức hệ thống			3.000.000
	- Kiểm thử giao diện			3.000.000
<b>7</b>	<b>Kết thúc dự án</b>	<b>150.000</b>	<b>20</b>	<b>5.000.000</b>
	- Bàn giao sản phẩm			1.000.000
	- Làm tài liệu hướng dẫn sử dụng			750.000
	- Đào tạo			300.000
	- Bảo trì			2.000.000
<b>Tổng</b>				<b>77.500.000</b>

Bảng 2.2: Dự toán chi tiết dự án

#### 2.2.4.3. Kết luận

Dự án hoàn toàn khả thi, sẵn sàng để thực hiện vì hệ thống đạt được yêu cầu về các chỉ tiêu, chức năng đặt ra.

Chi phí phù hợp, thời gian thực hiện khả thi khi đem về nhiều lợi ích cho công ty và nhân viên.

## 2.2.5. Yêu cầu người dùng

### Yêu cầu hệ thống:

Web nhận diện văn bản tiếng việt sử dụng Tesseract là hệ thống tao mới, chuyển đổi, cập nhật thông tin của các loại dữ liệu một cách nhanh chóng và chính xác. Từ đó tạo ra những thuận lợi, hữu ích cho người sử dụng web.

### Yêu cầu chức năng:

- Đăng nhập, đăng xuất tài khoản có sử dụng phân quyền giúp cho người dùng cũng như người quản trị có các quyền riêng của mình.
- Quản lý đưa ra các loại chuyển đổi: tức là các loại convert ví dụ như từ hình ảnh với đầu vào là các file ảnh như file .jpg, .png, ... chuyển đổi đầu ra sẽ là một file txt hay chuyển đổi từ dạng file .pdf sang file txt. Ngoài ra quản trị còn có chức năng thêm, sửa, xóa các loại chuyển đổi này.
- Quản lý hóa đơn: cung cấp cho admin việc quản lý hóa đơn bao gồm các chức năng như thêm, sửa, xóa thông tin hóa đơn khi khách hàng đăng ký dịch vụ.
- Quản lý thông tin các file người dùng: Kiểm soát được số lượng file mà người dùng đã sử dụng, ...

### Yêu cầu phi chức năng:

- Hệ thống có chức năng bảo mật và phân quyền.
- Người sử dụng phải đăng ký tài khoản và mật khẩu để có thể sử dụng các chức năng nâng cao của hệ thống.

### Yêu cầu giao diện:

- Giao diện cần phải đơn giản, thân thiện với người dùng.
- Dễ sử dụng đối với các đối tượng người dùng, tiện lợi trong quản trị và dễ bảo trì.

### Yêu cầu về tính hữu dụng của hệ thống:

- Giúp cho công ty, doanh nghiệp, trường học, ... kiểm soát được chất lượng lưu trữ cơ sở dữ liệu tránh những sai sót không đáng có xảy ra.

## 2.2.6. Kế hoạch dự án

### 2.2.6.1. Phân công nhiệm vụ

STT	Nhiệm vụ	Người chịu trách nhiệm	Ngày bắt đầu	Ngày kết thúc	Ghi chú
1	Team Leader	Trần Lưu Tú	04/04/2022	20/06/2022	
2	Nghiệp vụ viên	Nguyễn Huy Hoàng	04/04/2022	20/06/2022	
3	Phân tích viên	Nguyễn Văn Nam	04/04/2022	20/06/2022	
4	Lập trình viên	Nguyễn Văn Nam Trần Lưu Tú	04/04/2022	20/06/2022	

Bảng 2.3: Phân công nhiệm vụ

### 2.2.6.2. Kế hoạch dự án

STT	Công việc chính	Người chịu trách nhiệm	Ngày bắt đầu	Ngày kết thúc
1	Đề cương dự án	Nguyễn Văn Nam	04/04/2022	14/04/2022
2	Hồ sơ dự án khả thi	Trần Lưu Tú	10/04/2022	17/04/2022
3	Yêu cầu người dùng	Nguyễn Huy Hoàng	11/04/2022	13/04/2022
4	Kế hoạch dự án	Trần Lưu Tú	10/04/2022	20/05/2022
5	Tập yêu cầu và phiếu yêu cầu	Nguyễn Văn Nam	14/04/2022	20/04/2022
6	Tài liệu phân tích thiết kế	Nguyễn Huy Hoàng Nguyễn Văn Nam	17/04/2022	21/04/2022
7	Kế hoạch kiểm thử và tập test case	Nguyễn Huy Hoàng	01/06/2022	10/06/2022
7	Quản lý tiến trình phần mềm	Trần Lưu Tú	08/05/2022	12/05/2022
8	Tài liệu kết thúc dự án	Trần Lưu Tú	12/05/2022	14/05/2022
9	Tài liệu hướng dẫn	Trần Lưu Tú Nguyễn Văn Nam	14/05/2022	20/05/2022

Bảng 2.4: Kế hoạch dự án

### 2.2.7. Phân tích hệ thống

Web nhận diện tiếng việt sử dụng công nghệ Tesseract bao gồm các thành phần cơ bản như sau:

- Người dùng đăng nhập vào hệ thống để sử dụng các chức năng theo đúng quyền hạn của mình (sau đó đăng xuất nếu muốn)
- Người dùng sử dụng chức năng chuyển đổi các loại dữ liệu cụ thể là chuyển đổi dữ liệu từ dạng file ảnh hoặc file pdf sang dạng dữ liệu text.
- Người dùng nếu là admin thì có thể quản lý các file dữ liệu mà user đã sử dụng để chuyển đổi dữ liệu.
- Hệ thống sẽ trả về kết quả một file text cho người dùng. Sau đó người dùng có thể đánh giá thông qua chức năng reviews, ...
- Admin sử dụng chức năng quản lý hóa đơn để quản lý các user đã đăng ký sử dụng web.

Chi tiết cho từng modules trong hệ thống:

- Module đăng nhập, đăng xuất:
  - Đầu vào: thông tin người dùng bao gồm username (email) và password.
  - Chức năng: dùng để đăng nhập vào hệ thống sau đó sử dụng các chức năng như chuyển đổi dữ liệu từ định dạng file pdf hay định dạng file ảnh sang định dạng file text.
  - Đầu ra: Đăng nhập được vào hệ thống và sử dụng các chức năng tùy vào quyền hạn của người dùng.
- Module chuyển đổi các loại dữ liệu:
  - Đầu vào: là file có định dạng dưới dạng ảnh hoặc dạng file PDF.
  - Chức năng: Chuyển đổi các dạng dữ liệu từ file ảnh hay định dạng file pdf sang định dạng file text, giúp cho người dùng có thể tùy ý sử dụng file text đó để chỉnh sửa nội dung, lưu trữ thông tin tránh bị hư hỏng, ...
  - Đầu ra: Định dạng file text sau khi đã được chuyển đổi.
- Module quản lý file người dùng:

- Chức năng: dùng để quản lý các file mà người dùng đã sử dụng để chuyển đổi thành dạng file text cũng như quản lý số lượng file mà người dùng đã sử dụng để có thể chuyển đổi từ định dạng dữ liệu ảnh hay định dạng dữ liệu dạng file PDF sang dữ liệu dạng text.
- Module quản lý hóa đơn:
  - Chức năng: dùng để quản lý các đăng ký mà người dùng user đã đăng ký web nhằm sử dụng web cho công việc. Admin có thể thêm, sửa, xóa các thông tin của hóa đơn.
- Module cơ sở dữ liệu:
  - Chức năng:
    - Tổng hợp dữ liệu sinh ra từ các module vận hành (operational module) và lưu trữ vào cơ sở dữ liệu. Đảm bảo dữ liệu được lưu trữ ít nhất trong vòng một tháng
    - Cung cấp các báo cáo, dữ liệu tổng hợp.
    - Các báo cáo như: số lượng file cũng như số lượt chuyển đổi của người dùng sử dụng trong một ngày, ...
    - Kết quả có thể xem trên giao diện web quản trị hoặc xuất qua API đến các hệ thống khác.
- Module quản trị hệ thống
  - Chức năng:
    - Hiển thị thông tin và cho phép quản trị viên có thể thao tác, chỉnh sửa thay đổi cấu hình của hệ thống, kiểm soát các trạng thái của người dùng và module phần mềm.
    - Hiển thị thống kê số lượng các file được sử dụng để chuyển đổi, số lượt đã chuyển đổi của user trong ngày cũng như thống kê được các tài khoản đã sử dụng web.
- Module Api tương tác với các hệ thống khác:

- Chức năng: Hệ thống giao tiếp với các hệ thống khác thông qua WebService/ REST.

### 2.2.8. Rủi ro dự án

#### 2.2.8.1. Rủi ro trong công nghệ

STT	Rủi ro	P	I	E	Hành động giảm thiểu rủi ro
1	Phiên bản phần cứng, phần mềm thay đổi trong quá trình thực hiện dự án.	5 3	3 2	15 6	Người mua (khách hàng) nên yêu cầu có được phiên bản mới nhất hoặc giải pháp trọn gói.
2	Giao diện tới các hệ thống nội bộ và hệ thống bên ngoài có thể không hoàn thành đúng hạn.	4 2	3 1	12 2	1. Bắt đầu sớm, quy định chuẩn kết nối. 2. Cập nhật hệ thống cũ lên chuẩn mới. 3. Làm việc ngay với các hệ thống bên ngoài để thống nhất chuẩn.
3	Triển khai thử nghiệm có thể không đầy đủ cho việc xác định đảm bảo hiệu năng khi triển khai diện rộng.	4 2	3 1	12 2	1. Chuẩn bị, kiểm tra và tuân theo kế hoạch thử nghiệm chi tiết, xác định những câu phần quan trọng và những nghiệp vụ thiết yếu bắt buộc phải thử nghiệm. 2. Chấp nhận thực tế việc triển khai thí điểm sẽ không hoàn hảo và chỉ ra những gì cần sửa trước khi triển khai diện rộng cũng như những việc cần điều chỉnh sau đó.
4	Biện pháp phòng ngừa thảm họa có thể không như mong đợi.	3 2	3 3	9 6	1. Chỉ ra yêu cầu rõ ràng và cụ thể. 2. Yêu cầu kịch bản kiểm tra hệ thống phòng ngừa thảm họa.

Bảng 2.5: Rủi ro trong công nghệ

### 2.2.8.2. Rủi ro do con người

STT	Rủi ro	P	I	E	Hành động giảm thiểu rủi ro
1	Nguy cơ thành viên giỏi sẽ rời khỏi dự án giữa chừng vì chênh lệch thu.	5 3	5 3	25 9	1. Tạo ra cơ hội thăng tiến, học tập tại nước ngoài. 2. Lắng nghe cán bộ, đánh giá và áp dụng cải tiến của họ vào công việc.
2	Người dùng có thẻ không muốn sử dụng từ chối hệ thống mới.	5 3	5 3	25 9	Thuyết phục, tập huấn, tuyển thành viên mới, định kỳ luân chuyển cán bộ giữa các phòng ban.
3	Người dùng có thẻ không biết cách sử dụng hệ thống mới.	5 2	5 3	25 6	1. Thực hiện đào tạo, hỗ trợ người sử dụng cung cấp kiến thức về hệ thống mới. 2. Biên soạn, tập hợp đầy đủ tài liệu. 3. Thường xuyên theo dõi, giám sát việc hiểu và chấp nhận hệ thống mới (qua các ứng dụng hỗ trợ, thực hiện khảo sát), điều chỉnh việc tập huấn dựa trên các kết quả đánh giá này.
4	Lãnh đạo ở các cơ quan trực thuộc có thẻ không ủng hộ hệ thống mới.	3 1	5 3	15 3	1. Lãnh đạo cấp cao quan tâm thường xuyên tới dự án, thường xuyên gửi những báo cáo tích cực về dự án cũng như mong muốn về dự án tới các cấp lãnh đạo phía dưới. 2. Lãnh đạo cấp cao làm việc với lãnh đạo chính quyền địa phương để đảm bảo truyền thông điệp về dự án mới, đảm bảo sự chỉ đạo cũng như hỗ trợ của chính

Bảng 2.6: Rủi ro do con người

### 2.2.8.3. Rủi ro trong quá trình nghiệp vụ

STT	Rủi ro	P	I	E	Hành động giảm thiểu rủi ro
1	Quy trình nghiệp vụ có thể thay đổi trong khi triển khai dự án CNTT.	5 3	5 3	25 9	<p>1. Ban quản lý cấp cao phải tham gia vào quá trình xây dựng chính sách để đảm bảo hiểu rõ được ảnh hưởng của bất kỳ thay đổi quan trọng nào.</p> <p>2. Xây dựng các thay đổi cần thiết vào hệ thống thông qua quy trình quản lý thay đổi đã được phê duyệt.</p>
2	Phạm vi và yêu cầu có thể tăng, dẫn đến thiếu kinh phí.	5 3	5 3	25 9	<p>1. Mọi thay đổi về phạm vi sẽ phải qua tiến trình phê duyệt sự thay đổi và phải đảm bảo phù hợp với kinh phí được phê duyệt trước đó.</p> <p>2. Đặt các mức ưu tiên cho yêu cầu nghiệp vụ và yêu cầu kỹ thuật vào tiến trình ra quyết định.</p>
3	Hệ thống có thể không đáp ứng nhu cầu hoặc mong đợi của người dùng.	3 1	4 2	12 2	<p>1. Đảm bảo nhân viên ở các cấp tham gia vào việc tái thiết kế quy trình nghiệp vụ và hệ thống.</p> <p>2. Huy động người dùng vào việc kiểm tra và,</p>
4	Có những yêu cầu nghiệp vụ đặc biệt mà dự án không thể đáp ứng.	3 2	3 2	9 4	Báo cáo lên ban quản lý dự án những vấn đề này để chỉ ra ảnh hưởng tới quy trình nghiệp vụ và chiến lược kinh doanh.

Bảng 2.7: Rủi ro trong quá trình nghiệp vụ

#### **2.2.8.4. Chi tiết các loại rủi ro**

##### **Hạn chế, giám sát và quản lý các rủi ro**

- **Rủi ro thứ nhất:** phiên bản phần cứng, phần mềm thay đổi trong quá trình thực hiện dự án.
  - Giảm nhẹ hậu quả: yêu cầu khách hàng, người sử dụng cập nhật, chọn phiên bản mới nhất để sử dụng hoặc chọn giải pháp trọn gói.
  - Giám sát: theo dõi sự thay đổi của công nghệ và áp dụng sự thay đổi đó vào dự án hoặc chọn công nghệ phù hợp nhất đối với dự án.
  - Quản lý: quản lý và cập nhật liên tục các mã nguồn công nghệ mới.
- **Rủi ro thứ hai:** giao diện tới các hệ thống nội bộ và hệ thống bên ngoài có thể không hoàn thành đúng hạn
  - Giảm nhẹ hậu quả:
    - Bắt đầu sớm, quy định chuẩn kết nối.
    - Cập nhật hệ thống cũ lên chuẩn mới.
    - Làm việc ngay với các hệ thống bên ngoài để thống nhất chuẩn.
  - Giám sát: theo dõi dự án để có thể nắm bắt kịp thời các quy trình tạo giao diện tới các hệ thống để hoàn thành đúng hạn.
  - Quản lý: quản lý quy trình kết nối của hệ thống nội bộ với hệ thống bên ngoài.
- **Rủi ro thứ ba:** triển khai thử nghiệm có thể không đầy đủ cho việc xác định đảm bảo hiệu năng khi triển khai diện rộng.
  - Giảm nhẹ hậu quả:
    - Chuẩn bị, kiểm tra và tuân theo kế hoạch thử nghiệm chi tiết, xác định những cầu phần quan trọng và những nghiệp vụ thiết yếu bắt buộc phải thử nghiệm.
    - Chấp nhận thực tế việc triển khai thí điểm sẽ không hoàn hảo và chỉ ra những gì cần sửa trước khi triển khai diện rộng cũng như những việc cần điều chỉnh sau đó.

- Giám sát: theo dõi quá trình thử nghiệm để đảm bảo đầy đủ hiệu năng cho đến khi được triển khai trên diện rộng, theo dõi sát sao quá trình triển khai diện rộng để xử lý kịp thời.
- Quản lý:
  - Quản lý quá trình thử nghiệm.
  - Quản lý quá trình thực hiện phần mềm để kiểm soát hiệu năng ở mức độ nhất cho đến khi được triển khai trên diện rộng.
- **Rủi ro thứ tư:** biện pháp phòng ngừa thảm họa có thể không như mong đợi xảy ra bất ngờ
  - Giảm nhẹ hậu quả:
    - Chỉ ra yêu cầu rõ ràng và cụ thể.
    - Yêu cầu kịch bản kiểm tra hệ thống phòng ngừa thảm họa xảy ra.
  - Giám sát:
    - Giám sát liên tục và kiểm tra hệ thống.
    - Giám sát kịch bản đã đề ra về các thảm họa trước đó.
  - Quản lý:
    - Quản lý các kịch bản phòng ngừa thảm họa.
    - Quản lý các biện pháp mới được thêm vào.

### **Rủi ro do con người**

- Rủi do thứ nhất: Nguy cơ nhân viên giỏi sẽ rời khỏi dự án giữa chừng vì chênh lệch thu.
  - Giảm nhẹ hậu quả:
    - Tạo ra cơ hội thăng tiến, học tập tại nước ngoài.
    - Lắng nghe cán bộ, đánh giá và áp dụng cải tiến của họ vào công việc.
    - Tìm nhân viên bổ sung vị trí.
  - Giám sát:
    - Theo dõi dự án để sắp xếp nhân viên lấp chỗ trống.

- Quá trình làm việc của nhân viên để thường phạt rõ ràng để nhân viên có mục tiêu làm việc.
  - Các ý kiến của nhân viên với dự án để có lợi nhất cho dự án và biểu dương những ý kiến hay.
- Quản lý: quản lý nhân viên trong quá trình làm việc để tránh việc thiếu nhân lực khi nhân viên dời khỏi dự án.
- Rủi ro thứ hai: Người dùng có thể không muốn sử dụng hoặc từ chối hệ thống mới.
  - Giảm nhẹ hậu quả: thuyết phục, tập huấn, tuyển nhân viên mới, định kỳ luân chuyển cán bộ giữa các phòng ban.
  - Giám sát: quá trình triển khai của nhân viên với người dùng.
  - Quản lý: phát triển của phần mềm và triển khai đến người dùng một cách hợp lý.
- Rủi ro thứ ba: Người dùng có thể không biết cách sử dụng hệ thống mới.
  - Giảm nhẹ hậu quả:
    - Thực hiện đào tạo và hỗ trợ người sử dụng, cung cấp kiến thức về hệ thống mới.
    - Biên soạn, tập hợp đầy đủ tài liệu tập huấn.
    - Thường xuyên theo dõi, giám sát việc hiểu và chấp nhận hệ thống mới (qua các ứng dụng hỗ trợ, thực hiện khảo sát), điều chỉnh việc tập huấn dựa trên các kết quả đánh giá này.
    - Truyền thông rõ ràng về hệ thống mới tới nhân viên, nhấn mạnh việc nhân viên sẽ tham gia triển khai thí điểm để đảm bảo hệ thống mới đáp ứng yêu cầu của họ.
  - Giám sát:
    - Quá trình viết tài liệu tập huấn để người dùng hiểu một cách dễ nhất về phần mềm.
    - Quá trình đào tạo và hỗ trợ người sử dụng của nhân viên.
    - Việc hiểu và chấp nhận hệ thống mới.

- Quản lý:
  - Viết tập tài liệu đào tạo để cung cấp cho người dùng.
  - Quá trình đào tạo sử dụng hệ thống với người dùng.
  - Quá trình đào tạo tập huấn và chấp nhận hệ thống của người dùng.
- Rủi ro thứ tư: Lãnh đạo ở các cơ quan trực thuộc có thể không ủng hộ hệ thống mới.
  - Giảm nhẹ hậu quả:
    - Lãnh đạo cấp cao quan tâm thường xuyên tới dự án, thường xuyên gửi những báo cáo tích cực về dự án cũng như mong muốn về dự án tới các cấp lãnh đạo phía dưới.
    - Lãnh đạo cấp cao làm việc với lãnh đạo chính quyền địa phương để đảm bảo truyền thông điệp về dự án mới, đảm bảo sự chỉ đạo cũng như hỗ trợ của chính quyền địa phương trong khi thực hiện.
  - Giám sát:
    - Quá trình làm việc của các lãnh đạo để có thể họp khi không thống nhất ý kiến với ban lãnh đạo trực thuộc.
    - Quá trình tuyên truyền dự án với các ban lãnh đạo trực thuộc địa phương.
  - Quản lý:
    - Báo cáo tích cực của lãnh đạo về dự án và mong muốn về dự án tới.
    - Truyền thông điệp về dự án mới, đảm bảo chỉ đạo và hỗ trợ của lãnh đạo trong khi thực hiện dự án.

### **Rủi ro trong quy trình nghiệp vụ:**

- Rủi ro thứ nhất: Quy trình nghiệp vụ có thể thay đổi trong khi triển khai dự án công nghệ thông tin.
  - Giảm nhẹ hậu quả:
    - Ban quản lý cấp cao phải tham gia vào quá trình xây dựng chính sách để đảm bảo hiểu rõ được ảnh hưởng của bất kỳ thay đổi quan trọng nào.
    - Xây dựng các thay đổi cần thiết vào hệ thống thông qua quy trình quản lý thay đổi đã được phê duyệt.

- Giám sát:
  - Quá trình xây dựng chính sách để nắm bắt rõ sự ảnh hưởng của thay đổi.
  - Quá trình phê duyệt của bất kỳ thay đổi nào.
- Quản lý:
  - Chính sách của bất kỳ thay đổi nào để xử lý kịp thời.
  - Các thay đổi cần thiết của hệ thống.
- Rủi ro thứ hai: Phạm vi và yêu cầu có thể tăng, dẫn đến thiếu kinh phí.
  - Giảm nhẹ hậu quả:
    - Mọi thay đổi về phạm vi sẽ phải qua tiến trình phê duyệt sự thay đổi và phải đảm bảo phù hợp với kinh phí được phê duyệt trước đó.
    - Đặt các mức ưu tiên cho yêu cầu nghiệp vụ và yêu cầu kỹ thuật vào tiến trình ra quyết định.
  - Giám sát:
    - Quy mô và phạm vi của dự án để đảm bảo kinh phí phù hợp.
    - Các mức ưu tiên của quá trình xây dựng dự án.
  - Quản lý:
    - Quy mô, kinh phí đã cho trước của dự án.
    - Yêu cầu nghiệp vụ và yêu cầu kỹ thuật.
- Rủi ro thứ ba: Hệ thống có thể không đáp ứng nhu cầu hoặc mong đợi của người dùng.
  - Giảm nhẹ hậu quả:
    - Đảm bảo nhân viên ở các cấp tham gia vào việc tái thiết kế quy trình nghiệp vụ và hệ thống.
    - Huy động người dùng vào việc kiểm tra.
    - Thủ nghiêm nhiều chức năng chính.
  - Giám sát:
    - Quá trình phát triển của các version, các kết quả test của dự án xem có được chấp nhận hay không.

- Quá trình triển khai phần mềm xem có trực trắc hay không.
- Quản lý:
  - Quản lý các phiên bản code và test.
  - Triển khai phần mềm.
- Rủi ro thứ tư: Có những yêu cầu nghiệp vụ đặc biệt mà dự án không thể đáp ứng.
  - Giảm nhẹ hậu quả: báo cáo lên ban quản lý dự án những vấn đề này để chỉ ra ảnh hưởng tới quy trình nghiệp vụ và chiến lược kinh doanh.
  - Giám sát:
    - Tập yêu cầu mà khách hàng đưa ra.
    - Thông nhất với khách hàng qua từng phiên bản code.
  - Quản lý: phiên bản code, test.

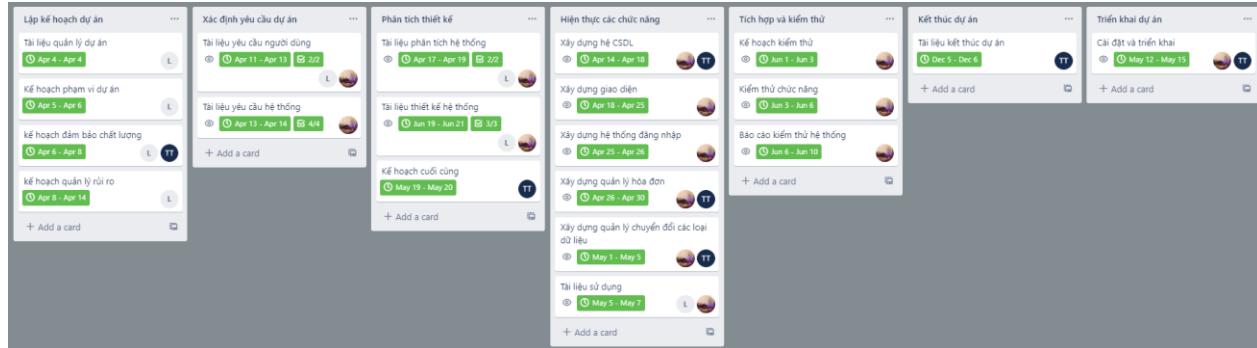
## 2.3. Công cụ quản lý dự án

### 2.3.1. Công cụ trello

- Tính năng nổi bật của Treollo:
  - Phân chia quản lý các công việc rõ ràng theo từng vị trí, từng tình trạng của đầu việc như: đang làm, đã làm xong, hủy bỏ, ...
  - Tổng hợp phân chia đầu việc tới từng người trong nhóm bao gồm mô tả công việc cụ thể, thời gian, tình trạng hoàn thành công việc.
  - Đây là công cụ hỗ trợ làm việc nhóm nên cho phép các thành viên trong nhóm có thể yêu cầu sự giúp đỡ của nhau khi thực hiện một nhiệm vụ nào đó.
  - Ưu điểm của Trello rất rõ ràng hướng đến sự đơn giản, linh hoạt và miễn phí. Với các tính năng cao cấp thì bạn mới phải trả phí. Tuy nhiên, với phần lớn công việc chúng ta chỉ cần dùng bản miễn phí đã quá đủ rồi.
- Các ứng dụng của Treollo trong quản lý dự án:
  - Treollo giúp bạn kiểm soát dự án một cách hiệu quả, từ các dự án về marketing, sale, thiết kế, ... tất cả đều được quản lý với một ứng dụng Treollo.
  - Treollo có thể giúp bạn lên kế hoạch cho bữa tiệc công ty, lên kế hoạch chuyển nhà, kiểm soát hàng hóa, ...

- Các công việc cần Checklist nhiều và nhiều người phối hợp làm việc chung như: tổ chức sự kiện, tiệc cưới, dã ngoại, ...

### 2.3.2. Áp dụng trong quản lý dự án



Hình 2.1: áp dụng trello trong quản lý dự án

### 2.4. Mô hình quản lý dự án

- Có rất nhiều mô hình quản lý dự án hiệu quả. Tuy nhiên, trong phạm vi môn thực tập, để quản lý một cách hiệu quả và thực hiện đúng tiến độ, em chọn quản lý dự án theo mô hình Scrum với mục đích:
  - Phân quyền người chịu trách nhiệm dự án.
  - theo dõi tiến độ và hỗ trợ các thành viên dự án giải quyết các vấn đề phát sinh.
- Thiết lập các mục tiêu ngắn hạn nhằm khuyến khích đẩy nhanh tiến độ.

### 2.5. Tự nhận xét và đánh giá

Thời gian	Công việc	Kết quả	Khó khăn
04/04/2022	Tìm hiểu Trello và mô hình phát triển phần mềm Scrum.	Hoàn thành	
05/04/2022	Phân tích thiết kế phần mềm.	Đang thực hiện	Các phân hệ đang trong quá trình hoàn thành
10/04/2022	Đánh giá các bước phát triển dự án.	Đang thực hiện	Một số phân hệ đã hoàn thành tương đối
11/04/2022	Tìm hiểu và lựa chọn thuật toán	Thực hiện và đang test thử	

20/04/2022	Xây dựng module login, tìm kiếm dữ liệu văn bản tiếng việt.	Đã hoàn thành	
28/04/2022	Test thử module login, và dữ liệu tìm được	Đã hoàn thành	
04/05/2022	Xây dựng module admin	Đang thực hiện	
10/05/2022	Xây dựng giao diện chuyển đổi các loại dữ liệu	Đang thực hiện	

Bảng 2.8: Nhận xét và đánh giá công việc

## CHƯƠNG 3: KẾT THÚC DỰ ÁN

### 3.1. Tài liệu kết thúc dự án

#### 3.1.1. Mô tả ngắn về dự án

Xây dựng web nhận diện văn bản tiếng việt sử dụng Tesseract để hoàn thành bản báo cáo thực tập môn học này.

#### 3.1.2. Lý do phát triển dự án

- Không đòi hỏi nhiều thời gian hay khách hàng, người dùng phải làm bất cứ điều gì khó khăn. Người dùng chỉ cần thao tác những thao tác vô cùng đơn giản mà lại đem lại lợi ích vô cùng to lớn.
- Kiểm soát theo thời gian thực
- Nâng cao mức độ an toàn cho dữ liệu nền, cũng như lưu trữ được các thông tin quan trọng của các loại tài liệu tại các doanh nghiệp, công ty, trường học, ... góp phần nào đó vào việc bảo đảm tài liệu sẽ được giữ lâu.

#### 3.1.3. Kết quả đạt được

- Về mặt thời gian: dự án hoàn thành chậm hơn tiến độ yêu cầu.
- Về mặt kỹ thuật: giúp cho việc chuyển đổi các định dạng dữ liệu như định dạng dữ liệu từ định dạng file ảnh hay định dạng file pdf sang định dạng file text một cách nhanh chóng, chính xác, giúp tiết kiệm thời gian cũng như chi phí, sức người.
- Dự án hoàn thành được các mục tiêu đề ra.

#### 3.1.4. Đánh giá về kết quả dự án

Tuy dự án hoàn thành chậm so với tiến độ những vẫn đáp ứng hoàn thành trong điều kiện nguồn lực cho phép. Tuy trong dự án có vấn đề phát sinh nhưng luôn được giải quyết kịp thời hướng đến kết quả nằm trong dự án.

#### 3.1.5. Một số vấn đề khi kết thúc dự án

- Sau khi bàn giao dự án và phần mềm đi vào hoạt động, toàn bộ nhóm chuyển sang trạng thái bảo trì.
- Kế hoạch phát triển dự án: khi khách hàng yêu cầu hoặc phần mềm bị lỗi thời thì sẽ liên hệ với khách hàng để đàm phán hợp đồng mới.

- Kế hoạch bảo trì:
  - Trường hợp 1: ngay khi nhận được phản hồi của khách hàng về phần mềm.
  - Trường hợp 2: nếu không có lỗi trong quá trình sử dụng phần mềm thì tiến hành bảo trì phần mềm 2 tháng / lần.

### 3.2. Kết thúc dự án

- Khái quát dự án:
  - Mục tiêu: dự án xây dựng web nhận diện văn bản tiếng việt sử dụng Tesseract.
  - Thành viên tham gia: 3 người.
  - Thời gian: khoảng 30 ngày.
  - Kinh phí: 150 triệu đồng.
- Tình trạng trước khi thực hiện dự án:
  - Các thành viên đồng ý thống nhất cùng nhau thực hiện dự án.
- Nội dung công việc của dự án:
  - Leader:
    - Xây dựng kế hoạch dự án, hồ sơ dự án khả thi.
    - Phân tích nhiệm vụ và deadline cho từng người.
    - Giám sát tiến độ công việc.
    - Nghiệp vụ viên.
    - Tham gia làm đề cương, yêu cầu người dùng.
    - Tập yêu cầu, phiếu yêu cầu.
    - Văn bản chiến lược kiểm thử.
    - Tài liệu đào tạo.
  - Phân tích viên:
    - Tham gia xây dựng dự án khả thi, Yêu cầu người dùng.
    - Tài liệu phân tích thiết kế...
    - Lập trình viên
    - Xây dựng văn bản ước lượng dự án, đưa ra con số thực tế trong phần mềm.
    - Xây dựng và quản lí code.

- Hướng dẫn cài đặt phần mềm....
- Tester:
  - Tham gia xây dựng văn bản yêu cầu người dùng.
  - Văn bản chiến lược kiểm thử.
  - Tập test case.
  - Quản lý tiến trình phần mềm.
  - Hướng dẫn sử dụng...
- Thành công và vần đề của dự án:
  - Với sự cố gắng miệt mài của tất cả các thành viên trong nhóm dự án đã hoàn thành với nhiều thành công như:
    - Các thành viên đều hoàn thành công việc của mình.
    - Xây dựng thành công website nhận diện văn bản tiếng việt bằng Tesseract.
  - Ngoài ra, còn có một số vấn đề phát sinh như sau:
    - Một số thành viên kinh nghiệm chưa cao, vẫn còn gặp phải một số vấn đề sai sót trong công việc của mình.
    - Một số công việc thực hiện sai tiến độ, ...
- Dự kiến sau dự án:
  - Thời gian bảo hành phần mềm:  
Trong vòng 3 năm khách hàng có thể liên hệ để được bảo trì và sửa chữa.  
Người phụ trách bảo trì: Nguyễn Văn Nam
  - Các module:  
Giữ lại các module để sử dụng cho các dự án về sau.  
Bổ sung các module mới nếu có phát sinh.
  - Thành viên:  
Không giải tán và chuyển sang giai đoạn bảo trì.  
Trau dồi thêm kiến thức, kinh nghiệm để phát huy cho các dự án sau.
- Kết luận:
  - Hệ thống sau khi test đã đưa vào vận hành và sử dụng trôi chảy.
  - Đã khắc phục những vấn đề và làm hài lòng khách hàng.
  - Đạt được mục đích và mục tiêu của dự án.
  - Thể hiện được tính thần và trách nhiệm của công ty với dự án.

### **3.3. Tài liệu đào tạo**

- Đối tượng được đào tạo: Nhân viên công ty.
- Yêu cầu đặt ra:
  - Đối với Admin quản trị:
    - Sử dụng thành thạo các chức năng của web.
    - Có khả năng đào tạo lại quá trình cho một Admin khác.
    - Thông hiểu các quy trình chuyển đổi dạng dữ liệu để hướng dẫn cho nhân viên.
    - Thành thạo các thao tác, chức năng.
    - Có thể xử lý các lỗi nhỏ.
    - Báo cho đội bảo trì nếu cần thiết.
  - Đối với nhân viên: thông hiểu về quá trình sử dụng website để chuyển đổi các định dạng dữ liệu.
- Lịch đào tạo: Ngay khi phần mềm được đưa vào thử nghiệm tại công ty.

## KẾT LUẬN

Quá trình nghiên cứu và thực hiện cài đặt chương trình theo sự khảo sát và yêu cầu của người dùng đã hoàn thành và đạt được một số kết quả cụ thể như sau:

- Hoàn thành đề cương của dự án.
- Hoàn thành hồ sơ dự án khả thi.
- Hoàn thành yêu cầu người dùng.
- Hoàn thành kế hoạch dự án.
- Hoàn thành tập yêu cầu và phiếu yêu cầu.
- Hoàn thành phân tích thiết kế cơ sở dữ liệu.
- Hoàn thành kế hoạch kiểm thử và test case.
- Hoàn thành tài liệu kết thúc dự án.

Một lần nữa, em xin chân thành cảm ơn các giảng viên khoa Công nghệ thông tin nói chung và thầy **Hoàng Thanh Tùng** nói riêng đã hỗ trợ, giúp đỡ em trong quá trình làm báo cáo thực tập lần này a.

## **TÀI LIỆU THAM KHẢO**

- [1] <https://github.com/tesseract-ocr/tessdata>
- [2] <https://tech.miichisoft.net/ocr-tesseract-js/>
- [3] [https://en.wikipedia.org/wiki/Tesseract\\_\(software\)](https://en.wikipedia.org/wiki/Tesseract_(software))
- [4] Video Node.js Express Project to Extract Text From Image Using Tesseract OCR Library in Browser Using JS
- [5]. Giáo trình quản lý dự án phần mềm, Vien CNTT – DHQG Hanoi
- [6]. Bài giảng: Quản lý dự án, Trường ĐH Điện Lực