BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**Đề tài**

**ỨNG DỤNG OPTICAL CHARACTER RECOGNITION  
VÀO HỆ THỐNG QUẢN LÝ CÔNG VĂN**

**Sinh viên: Trần Văn Hòa**

**Mã số: B1809127**

**Khóa: K44**

**Cần Thơ, 12/2022**

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**KHOA TRUYỀN THÔNG ĐA PHƯƠNG TIỆN**

**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**CHUYÊN NGÀNH TIN HỌC ỨNG DỤNG**

**Đề tài**

**ỨNG DỤNG OPTICAL CHARACTER RECOGNITION  
VÀO HỆ THỐNG QUẢN LÝ CÔNG VĂN**

**Người hướng dẫn**

**Ths Võ Hải Đăng**

**Sinh viên thực hiện**

**Trần Văn Hòa**

**Mã số: B1809127**

**Khóa: K44**

**Cần Thơ, 12/2022**

**LỜI CẢM ƠN**

# MỤC LỤC

[MỤC LỤC i](#_Toc119363172)

[DANH MỤC HÌNH ẢNH iv](#_Toc119363173)

[DANH MỤC BẢNG v](#_Toc119363174)

[DANH MỤC KÝ HIỆU CHỮ VIẾT TẮT vi](#_Toc119363175)

[TÓM TẮT vii](#_Toc119363176)

[ABSTRACT viii](#_Toc119363177)

[PHẦN GIỚI THIỆU 1](#_Toc119363178)

[ĐẶT VẤN ĐỀ 1](#_Toc119363179)

[LỊCH SỬ GIẢI QUYẾT VẤN ĐỀ 1](#_Toc119363180)

[MỤC TIÊU ĐỀ TÀI 2](#_Toc119363181)

[ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU 2](#_Toc119363182)

[NỘI DUNG NGHIÊN CỨU 2](#_Toc119363183)

[NHỮNG ĐÓNG GÓP CHÍNH CỦA ĐỀ TÀI 3](#_Toc119363184)

[BỐ CỤC CỦA QUYỂN LUẬN VĂN 3](#_Toc119363185)

[PHẦN NỘI DUNG 4](#_Toc119363186)

[CHƯƠNG 1: MÔ TẢ BÀI TOÁN 4](#_Toc119363187)

[1.1. MÔ TẢ CHI TIẾT BÀI TOÁN 4](#_Toc119363188)

[1.1.1. Optical Character Recognition 4](#_Toc119363189)

[1.1.2. Thuật toán cây quyết định 4](#_Toc119363190)

[1.1.3. Văn bản hành chính 4](#_Toc119363191)

[1.1.4. Quy trình quản lý công văn 6](#_Toc119363192)

[1.2. PHÂN TÍCH VÀ ĐÁNH GIÁ CÁC GIẢI PHÁP CÓ LIÊN QUAN 6](#_Toc119363193)

[1.3. TIẾP CẬN GIẢI QUYẾT VẤN ĐỀ, CHỌN GIẢI PHÁP 6](#_Toc119363194)

[CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT THUẬT TOÁN 7](#_Toc119363195)

[2.1. GIẢI PHÁP PHÂN TÍCH VÀ THIẾT KẾ MÔ HÌNH 7](#_Toc119363196)

[2.1.1. Sơ đồ Use Case 7](#_Toc119363197)

[2.1.2. Sơ đồ ER 7](#_Toc119363198)

[2.1.3. Sơ đồ Logical 7](#_Toc119363199)

[2.1.4. Sơ đồ phân rã chức năng 7](#_Toc119363200)

[2.1.5. Sơ đồ hoạt động 7](#_Toc119363201)

[2.1.5.1. Sơ đồ hoạt động chức năng đăng nhập 7](#_Toc119363202)

[2.1.5.2. Sơ đồ hoạt động chức năng thêm công văn đến 7](#_Toc119363203)

[2.1.5.3. Sơ đồ hoạt động chức năng sửa công văn đến 7](#_Toc119363204)

[2.1.5.4. Sơ đồ hoạt động chức năng phê duyệt công văn đến 7](#_Toc119363205)

[2.1.5.5. Sơ đồ hoạt động chức năng phân công xử lý công văn đến 7](#_Toc119363206)

[2.1.5.6. Sơ đồ hoạt động chức năng xử lý công văn đến 7](#_Toc119363207)

[2.1.5.7. Sơ đồ hoạt động chức năng từ chối công văn đến 7](#_Toc119363208)

[2.1.5.8. Sơ đồ hoạt động chức năng xóa công văn đến 7](#_Toc119363209)

[2.1.5.9. Sơ đồ hoạt động thống kê công văn đến 7](#_Toc119363210)

[2.1.5.10. Sơ đồ hoạt động thêm cán bộ 7](#_Toc119363211)

[2.1.5.11. Sơ đồ hoạt động sửa cán bộ 7](#_Toc119363212)

[2.1.5.12. Sơ đồ hoạt động xóa cán bộ 7](#_Toc119363213)

[2.1.6. Sơ đồ tuần tự 8](#_Toc119363214)

[2.1.6.1. Sơ đồ tuần tự chức năng đăng nhập 8](#_Toc119363215)

[2.1.6.2. Sơ đồ tuần tự chức năng thêm công văn đến 8](#_Toc119363216)

[2.1.6.3. Sơ đồ tuần tự chức năng sửa công văn đến 8](#_Toc119363217)

[2.1.6.4. Sơ đồ tuần tự chức năng phê duyệt công văn đến 8](#_Toc119363218)

[2.1.6.5. Sơ đồ tuần tự chức năng phân công xử lý công văn đến 8](#_Toc119363219)

[2.1.6.6. Sơ đồ tuần tự chức năng xử lý công văn đến 8](#_Toc119363220)

[2.1.6.7. Sơ đồ tuần tự chức năng từ chối công văn đến 8](#_Toc119363221)

[2.1.6.8. Sơ đồ tuần tự chức năng xóa công văn đến 8](#_Toc119363222)

[2.1.6.9. Sơ đồ tuần tự thống kê công văn đến 8](#_Toc119363223)

[2.1.6.10. Sơ đồ tuần tự thêm cán bộ 8](#_Toc119363224)

[2.1.6.11. Sơ đồ tuần tự sửa cán bộ 8](#_Toc119363225)

[2.1.6.12. Sơ đồ tuần tự xóa cán bộ 8](#_Toc119363226)

[CHƯƠNG 3: KIỂM THỬ VÀ ĐÁNH GIÁ 8](#_Toc119363227)

[PHẦN KẾT LUẬN 9](#_Toc119363228)

[KẾT QUẢ ĐẠT ĐƯỢC 9](#_Toc119363229)

[HƯỚNG PHÁT TRIỂN 9](#_Toc119363230)

[TÀI LIỆU THAM KHẢO 10](#_Toc119363231)

[PHỤ LỤC I BIỂU MẪU THƯỜNG GẶP 11](#_Toc119363232)

[PHỤ LỤC II CÁC BẢNG DỮ LIỆU 12](#_Toc119363233)

# DANH MỤC HÌNH ẢNH

# DANH MỤC BẢNG

# DANH MỤC KÝ HIỆU CHỮ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **Từ khóa** | **Diễn giải** |
| 1 | OCR | Optical Character Recognition (Nhận dạng ký tự quang học) |
|  |  |  |

# TÓM TẮT

***Từ khóa: NodeJS, Express, ReactJS, nhận dạng ký tự quang học, cây quyết định, quản lý văn bản hành chính.***

# ABSTRACT

**Keyword: NodeJS, Express, ReactJS, optical character recognition.**

# PHẦN GIỚI THIỆU

## ĐẶT VẤN ĐỀ

Bài toán, vấn đề đặt ra đối với đề tài là gì?

Tại sao cần làm đề tài này?

Trong bất kì lĩnh vực nào thì công tác quản lý luôn giữ một vai trò vô cùng  
quan trọng, trong đó có công tác quản lý công văn. Việc quản lý công văn trong các đơn vị hành chính và doanh nghiệp luôn là vấn đề cấp thiết, đòi hỏi có sự  
đầu tư cả về nhân lực và trang thiết bị.

Bên cạnh đó Chính phủ cũng ban hành các văn bản nhằm khuyến kích việc chuyển đổi số trong công tác quản lý công văn như: Nghị định số 30/2020/NĐ-CP ngày 05/3/2020 của Chính phủ về công tác văn thư có quy định về hệ thống quản lý tài liệu điện tử; Nghị quyết số 26/NQ-CP ngày 15/4/2015 của Chính phủ ban hành chương trình hành động của Chính phù thực hiện nghị quyết số 36-NQ/TW ngày 01/7/2014 của Bộ Chính trị Ban Chấp hành Trung ương Đảng về đầy mạnh ứng dụng, phát triển công nghiệp thông tin đáp ứng yêu cầu phát triển bền vững và hội nhập quốc tế;

Trước đây, việc quản lý công văn, giấy tờ thường theo quy cách truyền  
thống là lưu trữ trên giấy tờ, trong các cặp, tủ hồ sơ nên gây ra không ít phiền phức  
trong việc tra cứu, lưu trữ và bảo quản. Gần đây việc áp dụng các hệ thống quản lý vào việc quản lý công văn cũng đã dược sử dụng nhưng đa phần đều phải nhập liệu truyền thống. Do đó cần có một biện pháp quản lý tối ưu hơn trong việc quản lý công văn trong các đơn vị hành chính và doanh nghiệp.

OCR và Data Mining phát triển

Chính vì lý do đó, em quyết định nghiên cứu xây dựng một hệ thống quản lý bằng công nghệ OCR kết hợp với các thuật toán trong Data Mining để giúp việc quản lý công văn diễn ra nhanh chóng, dễ dàng và ít sai sót hơn.

Vì vậy trong luận văn sẽ tập trung tìm hiểu các kỹ thuật, các công nghệ cần thiết để xây dựng hệ thống **Ứng dụng Optical Character Recognition vào hệ thống quản lý công văn** làm đề tài luận văn tốt nghiệp của mình.

## LỊCH SỬ GIẢI QUYẾT VẤN ĐỀ

Trong khoa.

Trong nước.

Ngoài nước.

Đề tài nào đã giải quyết, những vấn đề nào chưa giải quyết được?

## MỤC TIÊU ĐỀ TÀI

Những vấn đề trọng tâm.

Hệ thống được xây dựng với mục tiêu cung cấp đầy đủ các tính năng cho quy trình nhận/gửi công văn của một hệ thống quản lý công văn thông thường tích hợp thêm việc trích xuất và nhận dạng các thành phần của công văn:

* Văn thư dễ dàng tìm kiếm công văn theo các tiêu chí như, ngày ban hành, số ban hành, người ký, … Dễ dàng quản lý trạng thái xử lý công văn cũng như phân loại công văn.
* Hỗ trợ việc nhận dạng và trích xuất các thành phần trên công văn giúp cho việc nhập liệu nhanh chóng và ít sai sót hơn.
* Hỗ trợ việc báo cáo hoạt động quản lý một cách trực quan, rõ ràng, nhanh chóng.
* Hỗ trợ quản trị viên quản lý hệ thống tốt hơn, dễ dàng thiết lập các dữ liệu và việc phân quyền được thực hiện linh hoạt.
* Giao diện hài hòa, thân thiện và hiện đại giúp người dùng dễ dàng sử dụng.

## ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

**Đối tượng nghiên cứu** là một hệ thống quản lý công văn tích hợp các thuật toán Data Mining và công nghệ OCR.

**Phạm vi nghiên cứu** của đề tài bao gồm các tính năng thêm mới công văn, xử lý các nghiệp vụ quản lý như: phê duyệt, phân công xử lý, xử lý, ban hành, …. Đặt biệt là chức năng trích xuất và nhận dạng các thành phần của công văn.

## NỘI DUNG NGHIÊN CỨU

Nghiên cứu về quy trình quản lý công văn, nghiên cứu về công nghệ OCR, xử lý ảnh, các thuật toán về Data Mining.

**Về lý thuyết:**

* Tìm hiểu về Optical Character Recognition.
* Tìm hiểu về các phương pháp xử lý ảnh.
* Tìm hiểu về các thuật toán Data Mining.
* Tìm hiểu về quy trình quản lý công văn theo quy định của nhà nước.

**Về người dùng và chức năng:**

**Phương pháp thực hiện:**

## NHỮNG ĐÓNG GÓP CHÍNH CỦA ĐỀ TÀI

Kết quả chính đạt được

## BỐ CỤC CỦA QUYỂN LUẬN VĂN

Bố cục của luận văn bao gồm các thành phần như sau:

**Phần giới thiệu:** Giới thiệu tổng quan về luận văn, đặt vấn đề, lịch sử giải quyết vấn đề, mục tiêu đề tài, đối tượng và phạm vi nghiên cứu, nội dung nghiên cứu.

**Phần nội dung:** Nội dung gồm 3 chương chính:

**Chương 1:** Mô tả bài toán

**Chương 2:** Thiết kế và cài đặt thuật toán

**Chương 3:** Kiểm thử và đánh giá

**Phần kết luận:** Tổng hợp kết quả đạt được và đề xuất hướng phát triển trong tương lai

# PHẦN NỘI DUNG

## MÔ TẢ BÀI TOÁN

### MÔ TẢ CHI TIẾT BÀI TOÁN

- Mô tả chỉ tiết bài toán: Giúp người đọc hiểu rõ các chức năng/tính năng/đặc điêm của sản phâm/phân mềm/hệ thông/giải pháp là mục đích cân đạt được của đê tài. Nêu là phân mêm thì mô tả các chức năng trông đợi. Nêu là giải pháp thì mô tả những vân đê mà giải pháp sẽ xử lý, cải tiên, khăc phục ...

#### Optical Character Recognition

#### Thuật toán cây quyết định

#### Văn bản hành chính

**“Văn bản”** là thông tin thành văn được truyền đạt bằng ngôn ngữ hoặc ký hiệu, hình thành trong hoạt động của các cơ quan, tổ chức và được trình bày đúng thể thức, kỹ thuật theo quy định.

**“Văn bản hành chính”** là văn bản hình thành trong quá trình chỉ đạo, điều hành, giải quyết công việc của các cơ quan, tổ chức.

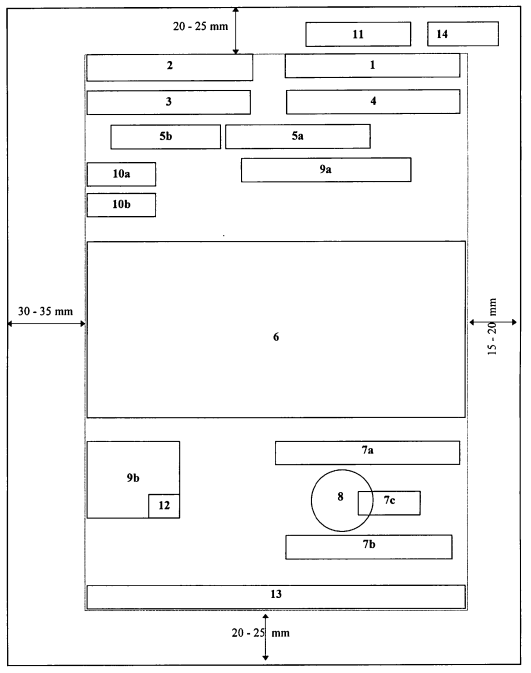
**“Văn bản điện tử”** là văn bản dưới dạng thông điệp dữ liệu được tạo lập hoặc được số hóa từ văn bản giấy và trình bày đúng thể thức, kỹ thuật, định dạng theo quy định.

**“Văn bản đi”** là tất cả các loại văn bản do cơ quan, tổ chức ban hành.

**“Văn bản đến”** là tất cả các loại văn bản do cơ quan, tổ chức nhận được từ cơ quan, tổ chức, cá nhân khác gửi đến.

**Sơ đồ bố trí các thành phần thể thức văn bản hành chính**

Trình bày theo chiều dài của khổ A4 (210 mm x 297 mm). Sơ đồ bố trí các thành phần thể thức văn bản hành chính theo hướng dẫn của **Phụ lục I thể thức kỹ thuật trình bày văn bản hành chính và bản sao văn bản** *(Kèm theo nghị định số 30/2020/NĐ-CP ngày 05 thánh 3 năm 2020 của Chính phủ)*



Hình 1.1 Sơ đồ bố trí các thành phần thể thức văn bản hành chính

**Vị trí trình bày các thành phần thể thức:**

|  |  |  |
| --- | --- | --- |
| **Ô số** | **:** | **Thành phần thể thức văn bản** |
| 1 | : | Quốc hiệu và Tiêu ngữ |
| 2 | : | Tên cơ quan, tổ chức ban hành văn bản |
| 3 | : | Số, ký hiệu của văn bản |
| 4 | : | Địa danh và thời gian ban hành văn bản |
| 5a | : | Tên loại và trích yếu nội dung văn bản |
| 5b | : | Trích yếu nội dung công văn |
| 6 | : | Trích yếu nội dung công văn |
| 7a, 7b, 7c | : | Chức vụ, họ tên và chữ ký của người có thẩm quyền |
| 8 | : | Dấu, Chữ ký số của cơ quan, tổ chức |
| 9a, 9b | : | Nơi nhận |
| 10a | : | Dấu chỉ độ mật |
| 10b | : | Dấu chỉ mức độ khẩn |
| 11 | : | Chỉ dẫn về phạm vi lưu hành |
| 12 | : | Ký hiệu người soạn thảo văn bản và số lượng bản phát hành |
| 13 | : | Địa chỉ cơ quan, tổ chức; thư điện tử; trang thông tin điện tử; số điện thoại; số Fax |
| 14 | : | Chữ ký số của cơ quan, tổ chức cho bản sao văn bản sang định dạng điện tử |

#### Quy trình quản lý công văn

### PHÂN TÍCH VÀ ĐÁNH GIÁ CÁC GIẢI PHÁP CÓ LIÊN QUAN

- Phân tích đánh giá các giải pháp có liên quan đến bài toán

### TIẾP CẬN GIẢI QUYẾT VẤN ĐỀ, CHỌN GIẢI PHÁP

- Tiếp cận giải quyết vấn đề, chọn lựa giải pháp

## THIẾT KẾ VÀ CÀI ĐẶT THUẬT TOÁN

Bao gồm thiết kế kiến trúc tổng thể của hệ thống, giải thích chức năng của từng thành phân trong hệ thông, các giải thuật xử lý của hệ thông hoặc của một thành phân hệ thông. thiêt kê cơ sở dữ liệu, ...

Mô tả cách thức cài đặt thiết kế bằng một ngôn ngữ lập trình cụ thể/hệ fflềuẳ Wlý1ýànhlphân cứng .... ị

### GIẢI PHÁP PHÂN TÍCH VÀ THIẾT KẾ MÔ HÌNH

#### Sơ đồ Use Case

#### Sơ đồ ER

Chưa đúng



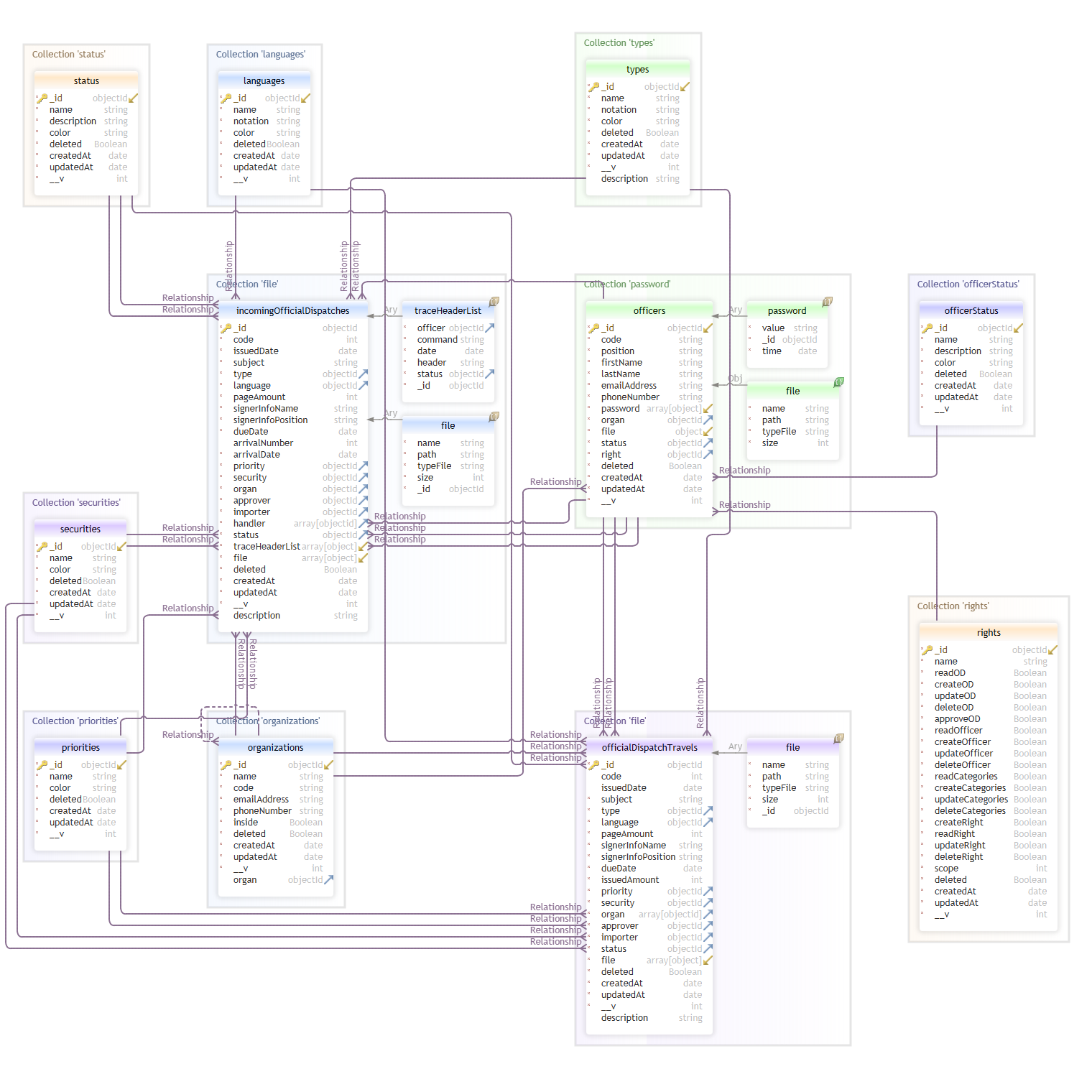
Hình 2.1 Sơ đồ ER

#### Sơ đồ LDM



Hình 2.2 Sơ đồ LDM

#### Sơ đồ LDM của MongoDB



Hình 2.3 Mô hình LDM

#### Sơ đồ phân rã chức năng



Hình 2.4 Sơ đồ phân rã chức năng

#### Sơ đồ hoạt động

##### Sơ đồ hoạt động chức năng đăng nhập

##### Sơ đồ hoạt động chức năng thêm công văn đến

##### Sơ đồ hoạt động chức năng sửa công văn đến

##### Sơ đồ hoạt động chức năng phê duyệt công văn đến

##### Sơ đồ hoạt động chức năng phân công xử lý công văn đến

##### Sơ đồ hoạt động chức năng xử lý công văn đến

##### Sơ đồ hoạt động chức năng từ chối công văn đến

##### Sơ đồ hoạt động chức năng xóa công văn đến

##### Sơ đồ hoạt động thống kê công văn đến

##### Sơ đồ hoạt động thêm cán bộ

##### Sơ đồ hoạt động sửa cán bộ

##### Sơ đồ hoạt động xóa cán bộ

#### Sơ đồ tuần tự

##### Sơ đồ tuần tự chức năng đăng nhập

##### Sơ đồ tuần tự chức năng thêm công văn đến

##### Sơ đồ tuần tự chức năng sửa công văn đến

##### Sơ đồ tuần tự chức năng phê duyệt công văn đến

##### Sơ đồ tuần tự chức năng phân công xử lý công văn đến

##### Sơ đồ tuần tự chức năng xử lý công văn đến

##### Sơ đồ tuần tự chức năng từ chối công văn đến

##### Sơ đồ tuần tự chức năng xóa công văn đến

##### Sơ đồ tuần tự thống kê công văn đến

##### Sơ đồ tuần tự thêm cán bộ

##### Sơ đồ tuần tự sửa cán bộ

##### Sơ đồ tuần tự xóa cán bộ

## KIỂM THỬ VÀ ĐÁNH GIÁ

Mô tả mục tiêu kiểm thử, kịch bản kiểm thử và kết quả kiểm thử: có chạy được hay không, chạy đúng không, đạt các mục tiêu đê ra hay không?

# PHẦN KẾT LUẬN

## KẾT QUẢ ĐẠT ĐƯỢC

## HƯỚNG PHÁT TRIỂN

# TÀI LIỆU THAM KHẢO

# PHỤ LỤC I BIỂU MẪU THƯỜNG GẶP

# PHỤ LỤC II CÁC BẢNG DỮ LIỆU